INNOVATIVE POLICIES TO MANAGE DEMAND IN SERVICE SYSTEMS WITH LIMITED CAPACITY

A Thesis
Presented to
The Academic Faculty

by

Naragain Phumchusri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2010

INNOVATIVE POLICIES TO MANAGE DEMAND IN SERVICE SYSTEMS WITH LIMITED CAPACITY

Approved by:

Professor Julie L. Swann, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Pinar Keskinocak
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Paul M. Griffin
Harold and Inge Marcus Department of Industrial and Manufacturing Engineering
Pennsylvania State University

Professor Bert Zwart
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Nicoleta Serban
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Anil Malhotra
Analytix Consulting, LLC

Date Approved: October 22, 2010

To Mom and Dad,

Ruangsri and Pimol Phumchusri,

for their greatest love and inspiration.

ACKNOWLEDGEMENTS

For the completion of this dissertation, I would like to thank the following people for their valuable contributions in making this dissertation what it is. First, I would like to express my appreciation and gratitude to my advisor, Dr. Julie Swann, for all her guidance and continuous support during my Ph.D. study. She has provided me not only constructive academic advices but also great encouragement in every step of this dissertation with patience. In each of our meetings, she kindly guided and persuaded me to advance my thoughts with greater creativity, and this motivated me to explore new ideas in improving our research. When I think about my journey at Georgia Tech, one of the most favorite times I recall will be working with Dr.Swann and I am very thankful.

I would like to thank my thesis committee: Prof. Pinar Keskinocak, Prof. Bert Zwart, Prof. Nicoleta Serban, Prof. Paul Griffin and Anil Malhotra for their valuable questions and comments. I truly appreciate their time and willingness to serve on my committee, as well as all their helpful suggestions. My gratitude also goes to the ISyE faculty for providing me with the foundations for my research, and to the ISyE staffs for their helps during the past five years of my study. Also, I would like show my gratitude to my senior friend, Dr. Serhan Duran, for his suggestion and collaboration in Chapter 2: Anil Malhotra and Bruce McMullen for the expertise and insights from their revenue management consulting experiences. I am thankful for the opportunity to work with them.

I would like to thank P. Chung for his tremendous resource of encouragement that brings out the best in me. His support motivates me to move forward in each step of my Ph.D. study. I am also grateful to have amazing friends here in Atlanta, who always provide me warm friendship. My thanks goes to my senior friends, P. Lek and P. To, for their guidance and courage for this work: P' Chi for cooking delicious meals that assist my brain to work effectively: Crystal, Bird, Pete, Uttam, P. Tam, Pui, Tarik and Vishal for their friendship and support through the good times and through the challenging times.

My deepest gratitude goes to my Mom and Dad, who are the main inspiration for me to try my best in every step of my life. They are the first and most important teachers who always support and guide me. Every time we talk, they always tell me, "All you have is today, so just try your best in each of every day...day by day...and eventually the light will show up at the end of your study journey". I cannot wait to show them I have seen that light. From the bottom of my heart, thank you Mom and Dad, for being such a great inspiration and for everything...

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This dissertation presents innovative demand management techniques for service systems with limited resources. The first study analyzes demand management policies of animal shelters with limited Kennel space as a set of interacting stochastic queueing systems. In practice, there are two main policies being used, which we call "Kill" and "No-Kill" policies. In a "Kill" system, animals may be euthanized if a shelter is full. Many shelters have moved to a "No-Kill" policy, where they avoid killing for space and adopt other approaches to reduce supply and demand mismatch. Our goal is to provide insights on how No-Kill policies, such as coordination, adoption and neutering campaigns, help reduce the animals' killing rate so that the shelter management can choose the way to effectively solve their problems.

In the second part, we consider a topic of demand management for the Sports and Entertainment (S&E) industry, called "Scaling the house", i.e., how to divide seats into zones for different prices to maximize revenue across the venue. From the data obtained from several performance venues in the U.S., we find ticket demand is impacted by locations of seats as well as by price. We characterize closed-form solutions for the optimal two-dimensional zoning decision (with row and column cuts) and the one-dimensional decision (with row cuts), and explore when each model should be applied. The third study considers pricing as a tool to manage demand for the S&E tickets. We develop dynamic pricing with demand learning models where demand is also affected by time left until the show dates. Since the show's popularity is usually uncertain to the seller, we propose a method to learn the overall popularity via Bayesian updates. We perform computational experiments to understand properties of the model solutions and identify when demand learning is most beneficial.

CHAPTER I

INTRODUCTION

Many organizations traditionally emphasized the supply side of the profit equation such as making cost reduction a strategic priority. However, rapidly growing technology results in expanding market opportunities and this shift causes many companies to realize that focusing on supply alone is not enough. Demand management has been spotlighted and considered to be an effective tool to generate competitive advantages in the present economic environment. In this thesis, we consider a variety of innovative demand management policies, including partial coordination, segmentations based on behaviors, and learning when demand also changes over time, with the goal of improving a system's performance, capacity utilization and customers service.

In this work we focus on the service industry, which is one of the major sectors in the US. Approximately 55% of the economic activity in the United States occurs in service industries as classified by the North American Industry Classification System (NAICS) [109]. The key characteristic of the service sector is that unlike most goods, in which firms can decide to sell before or after their manufacturing, service can only be sold prior to its production, not after [111]. This causes demand management of service systems to be complicated due to perishability issues. In addition, many service facilities may not have flexibility in changing the number of their resources in the short run. It is, therefore, important to manage demand so that they can effectively utilize their limited capacities.

In the first part of the dissertation (Chapter 2), we consider the problem of demand management in a public service system, namely, an animal shelter. City animal control organizations and individuals often bring stray animals to animal shelters (the city pound or individually operated entities). Individuals may also bring animals to the shelter for which they are unable to care. Approximately 5 million cats and dogs were euthanized in the shelter systems last year in the United States, which resulted in part due to the inadequate space in animal shelters [103]. For example, some shelters or city pounds kill animals after a short period of time when more animals arrive than the space available (denoted here as a Kill policy or a K system). Some shelters have moved to a No-Kill policy (called an NK system) where animals, waiting for adoptions, are not killed for space reasons, but as a result incoming animals may be turned away (e.g., left on the streets or sent to an alternate shelter) when there is insufficient space. The main goal of the No-Kill policy is to reduce killing rates with policies other than killing, e.g., encouraging coordination among shelters, promoting adoptions and emphasizing neutering practices.

We study such systems as a set of shelters as queues with diversion of animals when full. In the first part of Chapter 2, we consider the effects of the coordination policy and primarily focus on the rejecting (or killing) rate of animals, which is a key measurement of how well the shelter serves the overall community of strays and how many lives are lost due to limited capacity. We start with the basic model where a system consists of identical shelters that divert animals to another shelter when the first is full (i.e., partial coordination). We find that when each shelter has the same adoption rate distribution, the arrival rate of No-Kill system is identical to that of the M/M/mC/mC queueing system. However, when the adoption rate of each shelter in the system is different, we show that the killing rates of the coordinating No-Kill shelters and the M/M/mC/mC are not the same. For large systems, our results show that adding more coordinated shelters helps reduce the overall killing probability but the marginal improvement is decreasing. We also provide several insights on how to design different kinds of coordinating systems to achieve the highest improvement, including coordinating with neighbors or between particular pairs of shelters, as well as using one-way animals diversion when necessary.

In addition to shelter coordination, No-Kill shelters may encourage people to adopt animals by using adoption campaigns, e.g., using temporary storefronts, mobile units, or partnerships with retail operations. We identify how to obtain the optimal number of adoption campaigns to achieve the shelter's overall minimum cost when adoptions are concave and increasing with the number of the campaigns. Also, No-Kill shelters may attempt to reduce their demand (or the animals' incoming rate) by promoting a neutering policy in the community. When animals increase according to a type of growth function, we show that the shelter's total cost is first concave and then convex with the number of animals neutered. We give insights on which situation that each of the No-Kill policies can be most beneficial. For instance, our results suggest that an adoption campaign can be more attractive for a low campaign budget, while a neutering policy is more effective for a higher budget. Although our work in Chapter 2 is motivated by the animal shelters' policies, the context can be applied to other industries with limited servers, e.g., hospital operations, ambulances and internet servers.

In the second part of the thesis (Chapter 3), we study a demand management problem in the entertainment ticket industry that arose from discussions with decision makers in the performing arts organizations. Selling fixed perishable products, the entertainment ticket industry can potentially benefit from the revenue management idea but still has not received as much attention in the literature [41]. A key question for many venues is how to segment the venue by location and price (i.e., "Scaling the house"). We empirically explore transactional level data from 318 shows to observe how seating locations drive ticket demand within the venue. In addition to price, we find that distance from the stage and distance from a row's center also affect demand. We develop a two-dimensional zoning model for the optimal "Scaling the house" decisions. In addition, we present an alternative one-dimensional zoning model for the case where demand is not significantly sensitive to distance from the center. We show that the optimal seating row (to be priced at a higher price before switching to the next lower price) is the row whose expected revenue when charging a high price is equal to the expected revenue when charging a low price. We provide key comparative statics on how model parameters impact the optimal decisions. In robustness analysis, our results show that a risk-averse venue manager who tends to overestimate the sensitivities of demand may achieve higher revenue than a risk-taker manager who underestimates them. We also discuss important managerial insights on when it is most worthwhile to section seats into two-dimensional zones and how to match different types of shows (or different shapes of venues) to optimal seating segmentation.

While "Scaling the house" is extremely important, it is generally performed no more than once a year and organizations are now turning towards dynamic pricing as a tool to balance supply and demand dynamically as the season progresses. In the third part of the dissertation (Chapter 4), we consider a demand management problem via dynamic pricing for the Sports and Entertainment (S&E) industry. We develop dynamic pricing models for stochastic S&E demand in a discrete finite time setting, where demand depends not only on ticket prices but also on time left until the show dates. In our analysis of three venues with over 300 shows, demand of performances/arts tickets generally increases sharply in weeks before the show.

We assume the overall show popularity is uncertain to the seller, but this information can be learned via Bayesian updates as early sales are revealed. We consider two situations: (1) the seller can adjust ticket prices every period, or (2) the seller dynamically determines the optimal timing for one price change in the selling horizon and how much to change price. We perform computational experiments to understand properties of the model solutions and performances under different scenarios. Our results show that demand learning is most beneficial when the initial estimates are incorrect. In addition, it can be useful to consider demand pace when making price change decisions. Our results show that it is less necessary for the seller to vary price every period if a large amount of demand arrives close to the show dates. Our results also show that having flexibility to adjust price every period becomes more beneficial with the existence of demand learning.

Overall, in this thesis we consider innovative policies to manage demand in service systems, which are key components in our societies. We study a variety of decision levels from planning to real-time and consider policies including partial coordination, segmentation, and decision making incorporating demand learning. Although our work is motivated by problems specific to a particular type of organizations, we also contribute new results across organizations and sectors.

In each chapter of the thesis, we describe the contribution of our research in the literature review and present the main results as well as the key computational experiments. All proofs are provided in the Appendix.

CHAPTER II

ANALYSIS OF LIMITED CAPACITY SERVICE SYSTEMS BASED ON OPERATIONAL POLICIES OF ANIMAL SHELTERS

2.1 Introduction

In this chapter we study the operational policies (namely coordination, adoption and neutering policies) of animal shelters as a set of stochastic queueing systems. Although the specific problem we study is motivated by policies used to operate No-Kill animal shelters, similar systems can also be found in other systems of servers. We focus on the killing rate, which is defined as the rate of animals rejected from the capacitated system, and the total cost of shelters as the main measurements of this study.

"Approximately 5 million dogs and cats were killed last year in U.S. shelters. That is 13,800 every day or 575 dogs and cats killed every hour, 24 hours a day, seven days a week" [103]. The majority of these killings are due to space restrictions rather than diseases. Yet, this is happening in a society where a major league athlete who was involved in dog fights and deaths of a small number of animals was received with outrage and jail time. City animal control organizations and individuals often bring stray animals to animal shelters (the city pound or individually operated shelters). Individuals may also turn in animals for which they are unable to care. The vast majority of the deaths in these shelters are due to the limited capacity of animal shelters compared to the population of animals [81]. As a result of limited capacity, many shelters such as city pounds are operated with policies where they will kill animals after a period of time to gain more space for incoming animals.

Many people, especially those connected with shelters, find the killing rate unacceptable. In 1989, Ed Duvin wrote an article about humane societies and shelters, noting that the foundation of the system "Is the dark secret that it is, in part, little more than a vast killing machine?" [10]. As a result of the distaste for killing, many shelters have moved to a "No-Kill" policy. In this policy, the shelter does not euthanize any animals for space but only for reasons such as disease. When there is insufficient space, incoming animals may be turned away (e.g., left on the streets). To reduce the killing rate, shelters management attempts to reduce the killing rate by several methods such as 1) coordinating with each other, 2) increasing adoptions and 3) promoting animal neutering.

The current rate of about 5 million animals euthanized per year in shelters, though still unacceptable, is significantly reduced from about 17.8 million euthanized in 1985 [81]. The prevalence of No-Kill shelters may be one contributing factor to this reduction. The No-Kill policy offers much comfort to those who love animals, yet it has also generated controversy. For instance, one woman wrote to the Best Friends Animal Society "No-Kill shelters sound great. But where do all the other animals go?" [10]. Though the No-Kill movement has generated much discussion and controversy, little has been done to analyze its effects on the system. In this chapter, we analyze the main policies currently implemented in most No-Kill shelters: namely, coordination, adoption and neutering policies.

Basically, each shelter can be described with stochastic arrivals (of animals) into a limited capacity shelter with stochastic service or adoption rate environment. An animal is either accepted to wait for adoption or is turned away; thus there is no queueing within a shelter. When a coordination policy is implemented in the No-Kill system, incoming animals who arrive at a full shelter can be transferred to another shelter in the system if space is available. The adoption rates of the shelters may differ, depending on factors such as: location, physical conditions, management policies, etc.

If all shelters are full, we assume the incoming animals are rejected from the system. We refer to a set of: 1) coordinating No-Kill shelters as an NK system, and 2) uncoordinated shelters as a K system. In contrast to an NK system, if an incoming animal arrives to a full shelter in a K system, it is rejected without being diverted to a second shelter. We quantify the behavior of the system as measured by the killing (or rejection) rate of arriving animals. Of course, while animal deaths are important problems, the context of coordinated queueing systems translates across many industries with customers and servers, for example, hospitals, ambulances and internet servers.

For the issue of coordination among shelters, we first consider a basic model where a system consists of identical shelters. We show that when each shelter has the same mean adoption rate, the NK system gives the same rejecting rate as an $M/M/mC/mC$ system, even though the NK shelters only partially coordinate when a shelter is full. We find the killing rate of the NK system is less than that of the K system which has no coordination, but the rate of the reduced killing rate decreases as more shelters are included in the coordinating NK system. We find the killing rates of the NK and the $M/M/mC/mC$ system are not equal when the adoption rate of each shelter in the system is different. Intuitively, this is due to the ability to choose among shelters with different adoption rates when animals are diverted to another shelter in the NK system. For example, the time spent in the shelter before adoption (or inter-adoption time) can be reduced if an animal is diverted to a shelter with a better location (higher adoption rate). Otherwise, the time can be increased. Also, we consider a more general case when the mean adoption rate of each shelter is nonlinear, specifically an increasing concave function of the number of animals in the shelter. In this case, an NK system' killing rate is not necessary the same as an $M/M/mC/mC$'s rate. Yet, our numerical analysis shows that coordination still reduces the killing rate, although its benefit is small when the adoption rate is less sensitive to the number of animals in the shelter.

In practice, there may be different types of sheltered animals (e.g., puppies and older dogs), whose expected times in the shelter are not necessarily equal. We also take into account that animals arriving to the shelter can be divided into two different types: 1) those with shorter inter-adoption times (i.e., “good”), and 2) those with longer times (“bad”). We explore the impact of adoption and arrival rates of each type on the shelters' killing rate. When the shelter's mean adoption rates are different, we use computational experiments to explore a number of policy combinations (e.g., between a fast and a slow shelter versus between a pair of fast shelters or a pair of slow shelters). In some cases, it may be impractical to coordinate among all NK shelter, we examine the impact of coordinating only with shelters in neighborhoods and provide recommendations for effective policy selections.

In addition to stressing coordination, shelters with No-Kill policy attempt to reduce animal deaths by implementing adoption and neutering policies. The first policy can help reducing number of animals in the shelter, while the second can decrease the arrival rates. In typical service systems like call centers, reducing the arrival rates is not common since it means fewer customers (and usually less revenue). However, for non-profit organizations like animal shelters, having too many arrivals (e.g., animals) can cause a higher number of euthanized animals, so having fewer arrivals can benefit the system. In this chapter, we also explore the impacts of both adoption and neutering policies and consider which one is more effective than the other under different scenarios. When animals increase according to a type of growth function, and when a shelter promotes neutering, we show the shelter's total cost is first concave and then convex with the number of animals neutered. For the adoption policy, if adoptions are concave and increasing with the number of the campaigns, we find the shelter's total cost is convex in the number of campaigns used and there is an optimal number of campaigns for the minimum cost. We explore computational experiments to evaluate the total cost to compare adoption and neutering policies. We find that one policy can be more preferable than the other when campaign budgets are taken into consideration.

The rest of the chapter is organized as follows. In the next section we review the most pertinent literature. In Section 3 we describe the model and the assumptions for the coordination policy. We present the results when the shelters have equal mean adoption rates and when they have different adoption rates in Sections 3.1 and 3.2, respectively, and the insights from numerical experiments of this policy can be found in Section 3.3. Adoption and neutering campaigns are analyzed in Section 4. We conclude by summarizing our results and discussing the managerial insights derived from our analysis in Section 5. The proofs of all results are included in the Appendix.

2.2 Literature Review

This chapter draws on two main streams of literature; studies related to animal shelters' operations and queueing literature. Some papers have found the time to adoption can be different for animals with different characteristics (e.g., size and age). Posage et al. [118] point out that animals with gold or white coat colors, small size, and a history of an indoor environment spend less time in the shelters compared to others (those types of animals can be comparable to the "good" type in our study while others can be grouped as the "bad" type). Normando et al. [113] consider methods to increase successful adoption rates. The authors found that although a young age is the most important factor for a quick adoption, programs that include increased human interaction and special training for dogs with behavioral problems could aid in the successful adoptions. In addition, Diesel et al. [40] study how animal features (such as purebred status, size, sex, etc.) affect time to adoption. Their results show that the significant factors were breed, size, sex, neuter status on arrival at a shelter, color and age. Neither of all studies above consider operational policies of shelters; instead they primarily focus on the adoption effects of either the shelters' environment or animal characteristics. In contrast, our study focuses on how the performance of shelters can be improved by using different kinds of policies. To the best of our knowledge, we are the first to quantitatively analyze operational policies of animal shelters or study them as a stochastic system.

The second stream of work relevant to our study is the queueing literature. One area closely related to this chapter is the study of the Erlang loss formula. The Erlang loss formula (the blocking probability of the $M/G/C/C$ system) was first developed by A. K. Erlang and his colleagues (see [46]). Since then it has been widely studied and extensively applied by telecommunication researchers (see [136]). Important properties of the Erlang loss formula have been proved in several papers. Messerli [105] shows the rejecting probability of the $M/G/C/C$ system is a convex function of capacity C. Harel [64] proves that the Erlang loss formula is strictly decreasing and strictly convex in the service rate when the arrival rate and capacity C are fixed. In addition, the Erlang loss formula is shown to be independent of the service time distribution beyond its mean. In other words, the rejecting probability of the $M/G/C/C$ model is equal to that of the the $M/M/C/C$ system (see [124]). Other key properties of the Erlang loss formula can be found in [68] and [146]. While the papers above focus on the properties of the loss formula, we study various kinds of operational policies and consider impacts of different types of arrivals to the shelter queueing systems.

The type of queueing system closely related to our work is a parallel-server system, which is the most common model to support resource management of call centers (See [6], [53], [54], [59], [66], [71], [79], [114], [130], and [143]). In this chapter, we study animal shelters as systems with parallel servers with no waiting space. First, we focus on how the performance of No-Kill shelters can be improved by coordination among systems. Our study is, therefore, closely related to the capacity pooling and partitioning literatures. Smith and Whitt [131] mathematically prove that the probability of the $M/M/mC/mC$ system being full is strictly decreasing in m. They also show that combining the $M/M/C/C$ systems can lead to disadvantages when the service time distributions are different. While [131] study full coordination of resources, our work considers partial coordination, which only happens when one of the shelters is full. We prove that partial coordination gives the same rejection rate as full coordination when the service distribution is identical. Unlike the earlier study in [131], we assume the service (or adoption) time of each animal is based on the shelter to which the animal is transferred. We also show partial coordination of shelters in the NK system is not the same as the $M/M/mC/mC$ system when the service distribution of each shelter is different.

In other studies related to capacity pooling (e.g., [8]), the author presents performance bounds for the effectiveness of pooling situations and shows some trade-offs between capacity and utilization of the pooled versus the independent systems. Buzacott [23] studies the pooling of N servers in series when each server is dedicated to one task, versus N parallel servers when each server completes all the tasks. The author points out that the pooling system performs better when there is higher variability in the tasks. Sheikhzadeh et al. [129] consider machine sharing in manufacturing systems focusing on total flexibility and chaining. Benjaafar [9] study inventory pooling in systems with symmetric costs where supply lead times are created by the production system with finite capacity. In this chapter we include the exploration of performance improvement due to partial coordination among systems with different service distributions. We also computationally examine how to combine such systems to effectively reduce the rejection rate with a pooling policy.

Our study is also related to the problem of optimal control in queueing systems. Much of the literature in this area studies capacity planning; e.g., Kochel [78] considers how to optimally determine the number of servers and the waiting spaces for the M/M/C/K system. Borst et al. [20] focus on the M/M/C queues with the objective of minimizing both waiting and staffing costs where the waiting cost depends on the waiting time of customers. While those papers looked at capacity decisions, we focus on how shelters can improve their performance by using a variety of policies instead of changing their current capacities.

A number of studies consider the problem of deciding the optimal price for a queueing system. For instance, Mendelson [104] studies the problem of finding the optimal price for the computer service facilities in which the delay cost linearly increases with the waiting time. Dewan and Mendelson [39] consider the similar problem but allow a general delay cost function. Palaka et al. [115] study the M/M/1 system and decide optimal price and quoted lead time, where arrivals are assumed to be linear with price and lead time quoted. So and Song [133] extend the problem when demand is log-linear in price. Other papers on pricing decisions in queueing models can be found in [24], [63], [121] and [150]. In contrast to the literature where arrival rate is usually assumed to depend on prices, in our work we assume animal arrivals (or shelter demand) can be adjusted via neutering campaigns. We also focus on the overall cost, which is the killing cost plus campaign cost, rather than the system's profitability.

We build our work, first, on the assumption of identical customer (or animal) service (or adoption) time distribution. Then, to better represent animal shelter characteristics, we consider more than one type of animal. In our work, although the service time distribution is initially attached to each type of animal, it can be adjusted based on the shelter to which the animal is transferred. When shelters are located in different locations, each may have different traffic for people visiting for adoptions, or different animal preferences (e.g., big/small dogs in rural/urban settings). We provide an analysis of an adjusted model in which the adoption rate is a non-linear function of the number of animals in the shelters (which is not equivalent to the

$M/M/C/C$ system in this case). Moreover, we include the study on neutering and adoption campaigns to evaluate their impacts on shelters' performance. These ideas and assumptions distinguish the animal shelter systems from the previous studies of queueing systems, although our results hold across a range of problems.

2.3 Coordination Policy

We start our analysis of animal shelters with a coordination policy, namely, the NK policy where shelters coordinate with each other only when one of them is full. The first case we consider is when each shelter in the system has approximately the same mean adoption rate (Section 2.3.1) and the second case is when each shelter has a different mean adoption rate (Section 2.3.2). Computational experiments of both situations are presented in Section 2.3.3.

2.3.1 Identical Mean Adoption Rate

In this section, we consider systems with m identical shelters having capacity (the maximum number of animals allowed to be in a shelter) of C. Each space or cage for an animal is comparable to a server and each animal is considered as a customer in the system. Animals arrive to a shelter with Poisson arrival rate, λ. The overall service (or adoption) rate of the shelter, $i\mu$, is proportional to the current number of animals in shelter, i, since each animal's service or adoption time is considered to be exponentially distributed with mean of $\frac{1}{\mu}$. We assume when there are more animals in the shelter, the probability of people successfully finding the animal they would like to adopt also increases. In this section, each shelter in the K system is comparable to a $M/M/C/C$ system, with exponential arrival and service rates, and with capacity of C and no waiting spaces. Some of the assumptions will be relaxed for more general cases (e.g., when shelters have different types of animals, capacities and mean arrival rates, and when the overall service rate service is less than $i\mu$) in Section 2.3.2 and 2.3.3.

In a K system, there is no coordination between shelters; each shelter has an independent arrival rate λ. Once a shelter is at full capacity, e.g., all cages are occupied, then when another animal arrives, one is rejected and killed. Let m denote the number of shelters in a system. The states and transition rates between states of the K system when $m = 2$ are shown in Figure 1. We define the killing probability as P_C^K, which is the total long-run rejecting probability of each independent shelter (or the long-run probability of a shelter being full). From Gross and Harris [60], we know that

$$P_C^K = \frac{\frac{(\frac{\lambda}{\mu})^C}{C!}}{\sum_{i=0}^{C} \frac{(\frac{\lambda}{\mu})^i}{i!}}.$$

Since each shelter has arrival rate λ, the long-run rejection rate is λP_C^K. Hence, for a K system with m shelters, the total killing rate is $m\lambda P_C^K$. The notation is summarized in Table 1 for ease of reference.

Table 1: Notation for coordination policy

C	Capacity of each shelter
λ	Animal arrival rate to each shelter
μ	Adoption rate of each animal in the shelter
m	The number of shelters in each system
P_C^K	Killing probability of the NK system consisting of m shelters with capacity C
P_{mC}^{NK}	Killing probability of a K shelter with capacity C

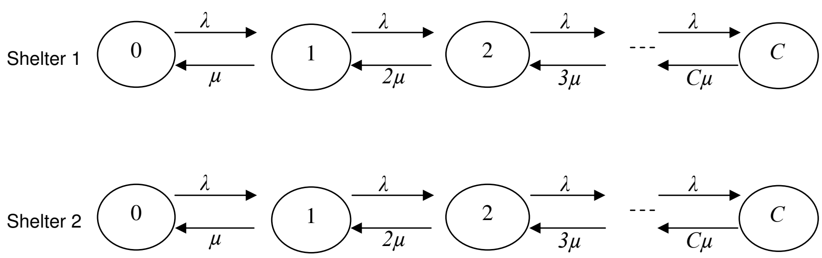

Figure 1: The transition diagram for a K system with two shelters.

If there is coordination among shelters in the system, An arriving animal can be transferred to another shelter when one of shelters is full. When there is no available shelter in the NK system, the arriving animals will be rejected. This is equivalent to being left on the street or being sent to a pound, in which case the animals' expected life time is quite short. Thus, in this section where each shelter is considered to be an $M/M/C/C$ system, the NK can be called a partial coordinated $M/M/C/C$ system since the coordination occurs only when one of them is full. The killing probability of the entire system is defined by P_{mC}^{NK}, which is the long-run probability of m shelters being full and the arriving animals cannot enter the NK system. The total arrival rate of an NK system with m shelters is $m\lambda$ since the system's inter-arrival time is the minimum of m exponential distributions with rate λ. Therefore, the killing rate of the entire NK system is $m\lambda P_{mC}^{NK}$. States and transition rates of the NK system with two shelters when $C = 3$ are shown in Figure 2. For $m = 2$, let (x, y) represent the state of the entire system where x and y are the number of animals in the first and the second shelters, respectively. From Figure 2, when one shelter is full, the arrival rate to the other shelter is doubled. For instance, the arrival rate from state (0,3) to state (1,3) is 2λ because the second shelter is already full, and arrivals to the second shelter are diverted to the first shelter.

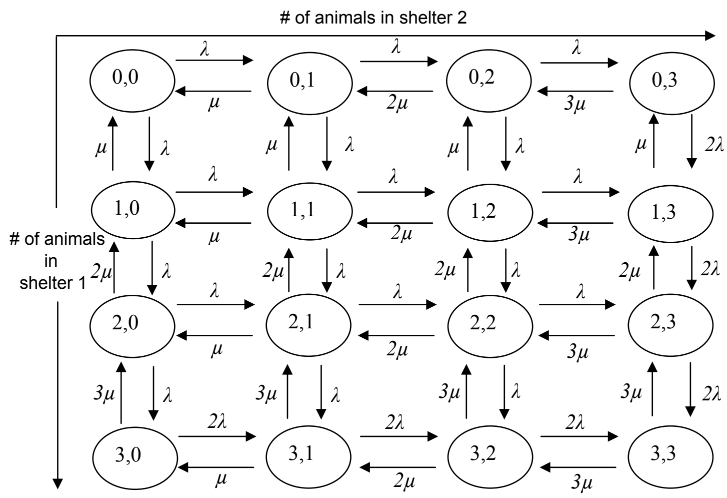

Figure 2: The transition diagram for an NK system with two shelters when $C = 3$

Theorem 1. *Given that the shelter capacity (C), arrival rate (λ) and the adoption rate (μ) are the same for each shelter in a system, the total killing rate of the system with m shelters under the NK policy is equal to the killing rate of an $M/M/mC/mC$ system with the total arrival rate $m\lambda$.*

The proofs of all the results are included in the Appendix. Theorem 1 implies that even though the nature of the NK and the $M/M/mC/mC$ system is different (NK only coordinates when necessary), both systems give an equal rejection probability in the long run. This is interesting since the NK system can be more decentralized and may be less expensive, if coordination is costly. We next compare the K system and the NK system based on their killing rates in the long run.

Corollary 1. *Given the shelter capacity (C), arrival rate (λ) and the adoption rate (μ) are the same for each shelter in a system, the total killing rate of the system with m shelters under the NK policy is less than that under the K policy.*

Due to Corollary 1, for systems with identical shelters and no coordination cost, the NK system is preferable. So far, we have discussed the situation when each shelter has one type of animal with the same adoption rate, μ. Next, let us consider what happens if there is more than one type of animal, which might be more realistic for most animal shelters. We will show the results we have proved are still valid in this setting.

Let animals arriving to the shelter belong to one of the two types, "good" or "bad", with Poisson arrival rates of λ_1 and λ_2, respectively. The inter-adoption times, the time animals spend in the shelter, are associated with their types and are exponentially distributed with mean, $\frac{1}{\mu_1}$ and $\frac{1}{\mu_2}$ (where $\mu_2 = k\mu_1$ and $0 < k < 1$) for good and bad type, respectively. We assume the bad-type animals, on average, take a longer time to be adopted than the good-type ones. Note that the model can be easily extended to consider more than two types of animals.

Each shelter having two types of arrivals in the K system is comparable to an $M/G/C/C$ system (no waiting space) with resultant Poisson arrivals rate of $\lambda_1 + \lambda_2$ and the associated service (adoption) time with mean, $\frac{\frac{\lambda_1}{\mu_1}+\frac{\lambda_2}{k\mu_1}}{\lambda_1+\lambda_2}$. Hence the mean adoption rate for each shelter is equal to $\nu = \frac{\lambda_1+\lambda_2}{\frac{\lambda_1}{\mu_1}+\frac{\lambda_2}{k\mu_1}}$. Note that the overall service time of a shelter is not exponentially distributed unless the service times of the good and bad type are equal (which is not true in this case). Let $P^K_{C,2-type}$ be the long-run killing probability of each shelter in the K system. Table 2 summarizes additional notation of this section for ease of reference.

Table 2: Additional notation for coordination policy

λ_1	Arrival rate of good-type animal to the shelter
λ_2	Arrival rate of bad-type animal to the shelter
μ_1	Adoption rate of each good-type animal in the shelter
μ_2	Adoption rate of each bad-type animal in the shelter
k	The ratio between the adoption rate of the bad-type animals and the adoption rate of the good-type animals ($\frac{\mu_2}{\mu_1}$)
r	The ratio between mean arrival rate and mean adoption rate of the shelter
SD^{NK}_{mC}	The successful diverting rate of a NK system consisting of m shelters
L^{NK}_{mC}	The long-run average number of animal in a NK system with m shelters
L^{K}_{mC}	The long-run average number of animal in a K system with m shelters
$P^K_{C,2-type}$	The killing probability of a K shelter with two types of animals

Since the $M/G/C/C$ system's steady-state probabilities are independent of the service distribution ([60]), the rejecting probability of the $M/G/C/C$ system is equivalent to that of the $M/M/C/C$ system and we have the following equation:

$$P^K_{C,2-type} = \frac{\frac{r^C}{C!}}{\sum_{i=0}^{C}\frac{r^i}{i!}}, \text{ where } r = \frac{\textit{mean arrival rate}}{\textit{mean adoption rate}} = \frac{(\lambda_1+\lambda_2)}{(\frac{\lambda_1+\lambda_2}{\frac{\lambda_1}{\mu_1}+\frac{\lambda_2}{k\mu_1}})} = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{k\mu_1}.$$

The next two properties address how the killing rate changes with arrival rates and adoption rates.

Property 1. *Under the K policy, the effect on the shelter's killing probability from reducing the arrival rate of bad-type animals is greater than the effect from reducing the arrival rate of good-type animals.*

Property 1 states that under the K policy, decreasing the arrival rate of bad-type animals yields a greater effect in reducing the killing probability than changing that of the good-type, independent of their proportion in the population. This is rational since when we have fewer animals spending a long time in the shelter (as compared to those with shorter time), it is likely that the shelter can service a higher number of animals in the long run.

In practice, a greater number of bad-type animals may be brought to the shelter, compared to good-type animals. Some shelters have a policy of not accepting animals of certain breeds (e.g., pit bulls) because they are harder to get adopted. However, this does not solve the cause of animals' overpopulation problem. Instead, a more effective method can be done by promoting regular public programs teaching the benefits of spaying and neutering. With cooperation of all shelters, the results can be substantial for successful reduction in shelters' killing numbers.

Property 2. *The killing probability of the animal shelter under the K policy decreases as the ratio between the adoption rate of the bad-type animals and the adoption rate of the good-type animals, k, increases ($0 < k < 1$).*

Property 2 shows the chance of killing animals in the shelter can be reduced by bringing the adoption rate of the bad-type animals closer to the adoption rate of the good-type animal (k close to 1). This can be done in practice by training bad-type animals in the shelters to be housebroken or obedient to increase their adoption rates.

As mentioned, it is shown in the literature ([124]) that the killing rate of a shelter with one type of animal (considered as an $M/M/C/C$ system) is equal to the killing rate of a shelter with two types of animals ($M/G/C/C$). Thus, the K systems with one type or two types of animals in each shelter also have the same killing rate as long as the mean adoption rates, the mean arrival rates, and the capacities of each shelter are equivalent. In the NK system, there is a partial coordination that occurs only when one of shelters is full. The NK system is, therefore, comparable to a partially coordinated $M/M/C/C$ system when there is one type of animal and comparable to a partial coordinated $M/G/C/C$ system when there are two types of animals arriving to the system. The next theorem explores an important characterization of the NK system.

Theorem 2. *Given that the shelter capacity, mean arrival rate and mean adoption rate are the same for each shelter in a system, the killing rate of the NK system with one type of animal (the coordinated $M/M/C/C$) is equal to the killing rate of the NK system with two types of animals (the coordinated $M/G/C/C$).*

Theorem 2 shows the NK system also has the same killing rate whether it has one or two types of animals. Hence, the results in Theorem 1 and Corollary 1 are also valid when we have two types of animals in each shelter. These results together show that when all m shelters in the system are identical, an NK system gives a lower killing rate than a K system does.

Next, we consider the following questions: 1) how can the average numbers of animals in the NK and the K system be compared?, and 2) how does the NK system perform in terms of its killing probability as we make more shelters available for coordination in the NK system? The first question is related to the ability of efficiently utilize the resources of each system, while the second question could occur when the No-Kill network gets larger by being coordinated with a higher number of No-Kill shelters. The next Lemma addresses these situations.

Lemma 1. *Given the shelter capacity (C), arrival rate ($\lambda_1 + \lambda_2$) and the adoption rate (ν) are the same for each shelter in a system,* *i) the long-run number of animals in the system with m shelters under the NK policy is greater than that in the system under the K policy.*

ii) the killing probability of the system with m shelters under the NK policy is decreasing in m.

From Lemma 1 (i), when all parameters are the same, there is a higher number of animals occupying the NK shelter's capacity, compared to the K shelters. Intuitively, it is because the NK system rejects fewer number of animals in the long-run. Although the arrival rate of each shelter in the NK system is similar to that in the K system, the NK shelters allow extra arrivals diverted from other NK shelters in the system when possible. As a result, the NK system utilizes its resources more efficiently since it serves more number of animals in the long run. It is also beneficial for adopters since there are more animals available in the NK shelters to be chosen.

Lemma 1 (ii) implies the NK systems consisting of a different number of shelters do not perform equally. As we increase the size of the system by adding more shelters to coordinate, the killing rate is decreased. So, having more centralization of resources decreases the killing probability. However, if we consider some costs of coordination, it might not be beneficial to do so, depending on how much a system improves compared to the coordination efforts. A further question is how the marginal benefit of adding more shelters to the NK changes with the number of shelters. We generate insights on this issue through computational experiments in Section 2.3.3.1.

We have seen that the NK system provides several advantages over the K system, our next question is; what situations (e.g., with respect to model parameters) are there for the NK system to be the most beneficial? We find that a possible way to quantify the benefits of the NK system is to measure the improvements of the NK system over the K system. This can also indicate when it is the most advantageous to switch from the K policy to the NK policy. We define this performance measure as the successful diverting rate, SD_{mC}^{NK}, i.e., the rate of animals successfully diverted to other available shelters in the system.

Observation 1. *For the NK system with m shelters, the successful diverting rate is given by:*

$$SD^{NK}_{mC} = m(\lambda_1+\lambda_2)(P^K_C - P^{NK}_{mC}) = m(\lambda_1+\lambda_2)\left(\frac{\frac{r^C}{C!}}{\sum_{i=0}^{C}\frac{r^i}{i!}} - \frac{\frac{(mr)^{(mC)}}{(mC)!}}{\sum_{i=0}^{mC}\frac{(mr)^i}{i!}}\right)$$

where $r = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{k\mu_1}$.

Observation 11 follows the idea that SD^{NK}_{mC} is the difference between the killing rate of the K system ($m(\lambda_1+\lambda_2)P^K_C$) and the NK system ($m(\lambda_1+\lambda_2)P^{NK}_{mC}$). It can be useful to see how the successful diverting rate changes with several key parameters to analyze the benefits of coordination, established in the next property.

Property 3.

i) $SD^{NK}_{mC} \to 0$ *as* $r \to 0$,

ii) $SD^{NK}_{mC} \to 0$ *as* $r \to \infty$,

iii) SD^{NK}_{mC} *is increasing in* m,

iv) SD^{NK}_{mC} *is increasing in* r *when* $P^K_C(\frac{C}{r} - 1 + P^K_C) > P^{NK}_{mC}(\frac{C}{r} - 1 + P^{NK}_{mC})$ *and decreasing in* r*, otherwise.*

From Property 3, when each shelter has either very low (case i) or very high (case ii) arrival rate compared to the adoption rate, the benefit from coordination (operating as the NK instead of the K system) does not help much to reduce the rejection rate. Methods other than coordination, such as increasing adoptions, might be more effective solutions for those cases. From Observation 2 (case iii), as the NK system gets larger (i.e., more shelters are coordinated) the successful diverting rate also gets larger, which is quite intuitive; while case iv gives the condition how the successful adoption rate is affected by the ratio of arrival and adoption rate of each NK shelter.

What we have discussed so far relates to the case where each shelter in the system is identical. In practice, shelters might differ in size or be operated in different environments; for instance, one may have a higher incoming rate due to a greater animal population. However, as long as the total arrival rates and the total capacities of the entire system are the same (but shelters in the same system are not necessarily identical), the killing rate under the NK policy is less than the rate under the K policy. The proof is very similar to the proofs of Theorem 1 and Corollary 1, therefore we omit it.

The advantage of the identical adoption rate assumption is that it allows us to directly compare the K with the NK system when the animals in the shelter are not identically adoptable, e.g., when there are more than one type of animal. We next explore a more general situation where each shelter in the system may have different mean adoption rates. This can happen if one shelter is located in a populated area with more people visiting to adopt animals; while the other shelter is not. It is interesting to see if the NK system still rejects (or kills) fewer animals compared to the K system.

2.3.2 Different Mean Adoption Rates

In this section, we consider a more general case when the shelters' mean adoption rates are different (e.g., the adoption rates of the each type of animal are not equal). We define the mean adoption rate of the good-type and the bad-type animals as $\mu_{1,1}$ and $\mu_{2,1}$ for the first shelter and $\mu_{1,2}$ and $\mu_{2,2}$ for the second shelter, respectively. When there is coordination between shelters, the adoption rates of animals can be adjusted. For example, if an animal from a shelter is diverted to a popular shelter with higher number of adopters (e.g., in a bigger city), it may take less time for him/her to get adopted. On the other hand, when an animal is diverted to a less popular shelter, it may take longer time until someone decides to adopt him/her. We assume the adoption rate of an animal will change, depending to which the shelter it is transferred.

The arrival rate for each type of animal is the same as the previous section, λ_1 for the good type and λ_2 for the bad-type animals. Let ν_1, and ν_2 be the mean adoption rates of the first and the second shelter, respectively. We have:

$$\nu_1 = \frac{\lambda_1+\lambda_2}{\frac{\lambda_1}{\mu_{1,1}}+\frac{\lambda_2}{\mu_{2,1}}} \text{ and } \nu_2 = \frac{\lambda_1+\lambda_2}{\frac{\lambda_1}{\mu_{1,2}}+\frac{\lambda_2}{\mu_{2,2}}}.$$

Observation 2. *Given that the mean adoption rate of each shelter in the system is not identical, the total killing rate of the NK system can be different than the rate of the $M/G/mC/mC$ system.*

We prove observation 13 through a counter example by showing that Theorem 1 cannot be applied when the shelters' adoption rates are not equal. In fact, there is no closed-form solution for the NK's killing probability in this case. To get further insights, we use computational experiments to show the performance of the K policy versus the NK policy by comparing the long-run rejecting probability of both systems, presented in the next section.

2.3.3 Numerical Examples and Insights

In this section, we use numerical examples to obtain some insights about the performance of the shelter policies and their relationships with problem parameters such as arrival rates, adoption rates and shelter capacities. Computational experiments also allow us to examine some systems that do not have a closed-form solution; e.g., the NK system with shelters having different mean adoption rates.

The numerical examples are divided into three parts. Section 2.3.3.1 is related to Section 2.3.1, where the mean adoption rates of the shelters are equal. Section

2.3.3.2 corresponds to Section 2.3.2 where the mean adoption rates of the shelters are different. If there is only one type of animal in the NK system (the coordinated $M/M/C/C$), the killing probability can be obtained by solving the balance equations for the system such as illustrated in Figure 2 where the adoption rates (μ) for shelter 1 and shelter 2 are different. If there are two types of animals in the NK system (the coordinated $M/G/C/C$), we use simulation to find the long-run killing probability in the numerical examples. It also allows us to obtain the rejecting probability of systems with a combination of K and NK policies; i.e., some shelters in the system use the K policy while others use the NK policy. Section 2.3.3.3 shows what happens if we alternatively consider the shelter whose the mean adoption rate is nonlinear, specifically an increasing concave function of the number of animals in the shelter. Managerial insights on coordination policy analysis are provided in Section 2.3.3.4.

2.3.3.1 Systems of shelters with equal mean adoption rates

We perform computational experiments for the case where each shelter has the same mean adoption rate in this section. In the first analysis, we consider how much the killing probability of the NK system is lower than that of the K system. Next, we explore if shelters cannot coordinate, how much the inter-adoption time of each individual shelter needs to be reduced to achieve the same benefit as performing coordination. Then, we examine the benefit gained of adding more shelters to the NK system.

Let us consider the first analysis. We proved in Section 2.3.1 that the NK gives a lower adoption rate than the K system does. However, it is interesting to quantify the difference. Consider systems with two shelters in which the capacity of each shelter is 40 and $k = 0.5$. The good-type and bad-type animals arrive with Poisson arrival rates, 2 and 3 per day, respectively, so the mean adoption time is equal to $\frac{\frac{2}{\mu_1}+\frac{3}{0.5\mu_1}}{2+3} = \frac{8}{5\mu_1}$. Figure 3 presents the killing probability of the described K and the

NK systems at different inter-adoption times ($\frac{1}{\mu_1}$) of the good-type animals on the x-axis.

As proven in Corollary 1, the NK's killing probability graph is below the K system's. According to Observation 1 (i and ii), as the inter-adoption time becomes very large ($r \to \infty$), or becomes very small ($r \to 0$), the killing probabilities of both system converge to each other. The explanation is; as r goes to infinity, the probability of both systems being full is very high. When animals arrive at a full shelter in a NK system, there is a very small probability of other shelters not being full. The rate of successful diverting, SD_{2C}^{NK}, is also very small. The NK system, therefore, rejects animals almost as often as the K system does. On the other hand, as r goes to zero, SD_{2C}^{NK} also converges to zero since each shelter is likely to be available at all times and diverting is unnecessary. It implies *the shelter does not benefit much from coordinating as an NK system. Especially when the existing killing rates of shelters are already high, shelter management should focus on decreasing the number of abandoned animals or increasing adoptions rather than using a coordination policy.*

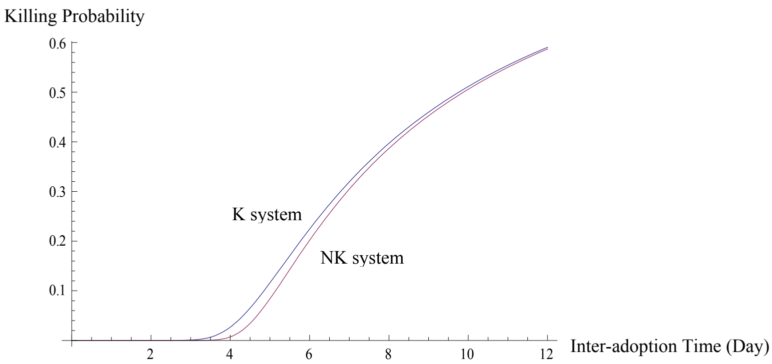

Figure 3: The killing probability of the NK and the K systems with two shelters when $C = 40$

To explore how increasing adoptions can improve the shelter system compared to using a coordination policy, we consider the following question: how much would the inter-adoption time of each animal in a K system need to be, in order to reduce the killing probability to the same as an NK system? We would like to see if shelters cannot (or do not want to) coordinate, how much their inter-adoption time need to be reduced to achieve the same benefit as performing coordination. An additional motivation for this question is: in practice, the inter-adoption time of the NK may be higher than the K system since the less desirable animals may be left in the NK shelters. Given all other parameters are the same, Table 3 presents the percent animal's adoption time (i.e., the time each animal spends in a shelter until being adopted) required for K system to be lower than the NK system's in order to achieve the same level of killing probability.

Table 3: Percent of adoption time required for the K system to be lower than the NK system in order for the K system to achieve the same level of killing probability as the NK system ($C = 40$)

Killing probability	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45
% decrease in adoption time	9.57	6.80	4.76	4.17	3.84	3.57	1.96	1.84	1.32

From Table 3, when the acceptable killing probability is at most 0.3, the K system will have the same killing probability as the NK system if the K system's adoption time is 3.57 percent lower than the that of the NK system. One possible way to improve the adoption time is: instead of rejecting every animal who arrives when the K shelters are full, one may replace the animal who has spent the longest time in the shelter by the arriving one. Furthermore, if the adoption time of the uncoordinated K system is reduced, as shown in Table 3, then its killing probability can be equal to the coordinated NK system. If the adoption time can be decreased by a greater percentage than the numbers shown, coordination might even be unnecessary.

From Lemma 1 (ii), the killing probability of the system with m shelters under the NK policy is decreasing in m. It is of interest to learn how much benefit is obtained by adding one more shelter to the NK system. This can also quantify whether a large coordinating NK system is beneficial. Consider an NK shelter where the adoption times of the good-type and bad-type animals are 24 and 36 days respectively. The good-type and bad-type animals arrive with Poisson arrival rate, 0.5 and 1 per day, respectively. Figure 4 shows the killing probabilities of the NK system consisting of 2,3,4,...,10 NK shelters described above.

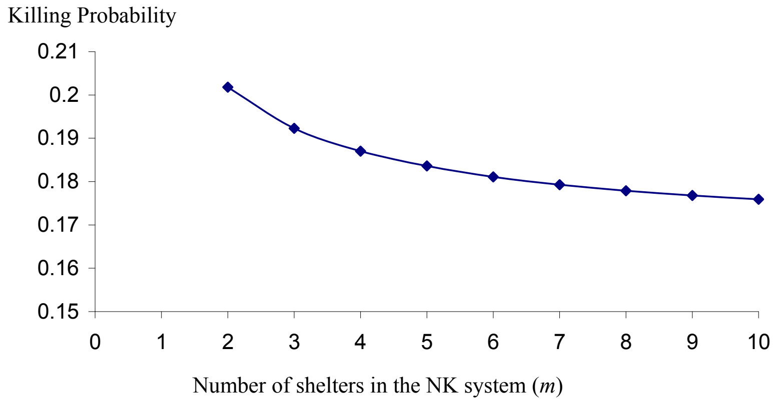

Figure 4: The killing probability of the NK system consisting of different numbers of shelters

Observation 3. *The benefit gained from adding one more shelter into the NK system decreases as the existing number of shelters in the system increases.*

Not surprisingly, this example illustrates that as we add one more shelter to coordinate in the NK system, the killing probability reduction is greater when the NK is small than when the NK system is large. Thus, if there is a fixed cost of coordination for every shelter added, increasing the size of the NK system to be too large might not be a good idea. We should, therefore, consider the trade-off between the benefit gained and the increased cost to find the optimal number of shelters for each NK system. For large systems, an alternative choice might be coordination among a subset of No-Kill shelters, in which we will explore in the next analysis.

In big cities, all No-Kill shelter may not be close to each other. Some shelters may be able to divert animals to others in their neighborhoods, but not to those too far away. Specifically, we consider No-Kill systems with shelters that coordinate with only shelters located next to them (defined as a "NK-Neighbor" system). To illustrate the structure of the described system, Figure 5 (left) depicts the coordination of the "NK-Neighbor" systems when the number of shelters (m) equals 4, 5, or 6 respectively, reading from top to bottom. The number of connections needed for the NK coordination (each shelter connects to all others) and the "NK-Neighbor" coordination is shown in Figure 5 (up, right). Figure 5 (down, right) presents the killing probabilities of the K system with no coordinations, the "NK-Neighbor", and the NK system, when there are 4 to 10 shelters in each system. In this example, C=40, and the good-type and bad-type animals arrive with Poisson arrival rate of 0.5 and 1 per day, respectively. We consider two scenarios; 1) when the adoption times of the good-type and bad-type animals are 24 and 36 days, and 2) when they are 20 and 30 days, respectively. Thus, the offered load (r) is equal to 0.5*24+1*36=48 in the first case and 0.5*20+1*30=40 in the second case.

From Figure 5 (down, right), the killing probability of the K system is the highest in both scenarios, as expected. An observation is that the killing probabilities of the "NK-Neighbor" and the NK system are much lower than the K system in the first case (r=40, with lower adoption time), compared to in the second case (r=48, with higher adoption time). With high r, the NK is not that much better than the K system, consistent with earlier insights. As the number of shelters increases from 4 to 10, the advantage of the NK increases over "the NK-neighbor". However, the marginal benefit of the NK over the "NK-neighbor" is much less than that of the NK over the K system. Moreover, from Figure 5 (up, right), as m increases, the NK system requires a much higher number of connections than the "NK-Neighbor" does. Consequently, *for a large No-kill system, it may not be necessary for shelters to coordinate with all others. Systems similar to the "NK-Neighbor" coordination can be an effective and practical option.*

So far, we have shown computational experiments of the situation where the mean

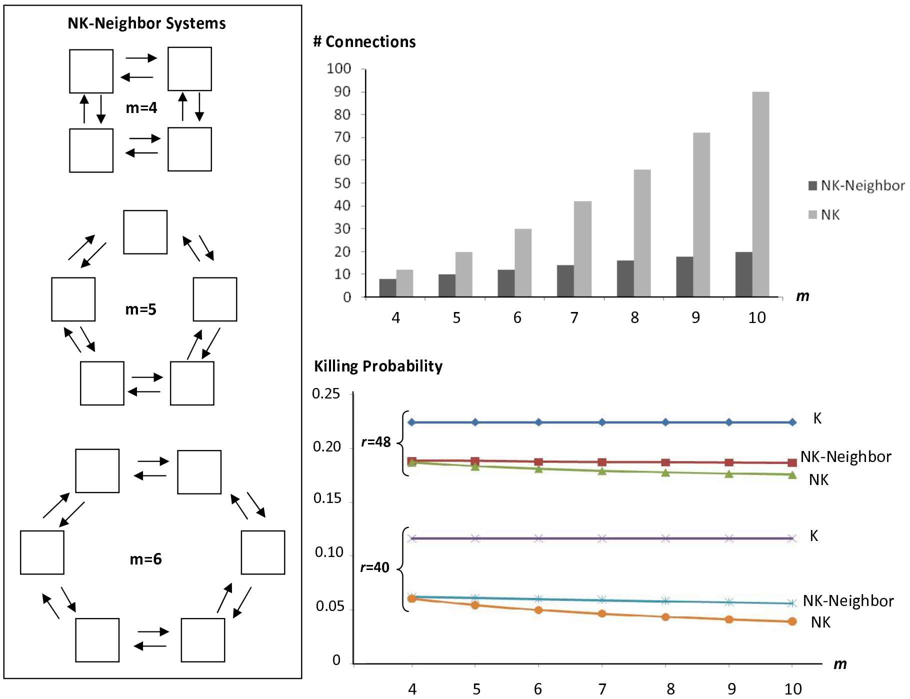

Figure 5: The "NK-Neighbor" systems when the numbers of shelters (m) are 4, 5 and 6 respectively (left); the number of connections required for the "NK-Neighbor" and the NK coordinations (up, right); and the killing probabilities of the K, the "NK-Neighbor" and the NK system (down, right)

adoption rate of each shelter is the same. Let us next consider what happens if this assumption is relaxed.

2.3.3.2 Systems of shelters with different mean adoption rates

As mentioned in Observation 3, when the mean adoption rate of each shelter in the system is not equal, the total killing rate of the NK system can be different than the rate of the $M/G/mC/mC$ system. In this section, we obtain some insights by exploring a variety of examples.

Consider systems of two shelters (C=30), where arrival rates to each shelter are 2 and 4 per day for the good and the bad types, respectively. The first shelter in the system has good-type adoption time, 6 days, and bad-type adoption time, 12 days. Thus, the mean adoption time for the first shelter is equal to $\frac{6*2+12*4}{2+4} = 10$ days

($\nu_1 = 0.1$ and $r = \frac{2+4}{0.1} = 60$). The killing probability of the first shelter is equal to $P_1 = \frac{\frac{(60)^{30}}{30!}}{\sum_{i=0}^{30} \frac{(60)^i}{i!}} \approx 0.515$. Figure 6 depicts the killing probability of the systems with two shelters, at different mean adoption rates of the second shelter (ν_2) on the x-axis.

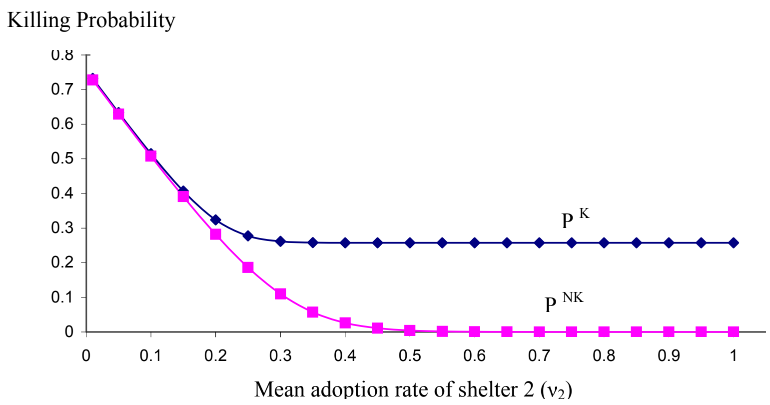

Figure 6: The killing probability of the system operated with the K and the NK policy when $\nu_1 = 0.1$ and $C = 30$

Observation 4.

i) *The killing probability of a NK system is less than that of the K system when each shelter in the system has different mean adoption rate.*

ii) *Given shelter 1's mean adoption rate is fixed, the difference between the killing probabilities of the K and the NK system is non-decreasing in the mean adoption rate of shelter 2.*

iii) *As* $\nu_2 \to \infty$, $P^{NK} \to 0$; *while* $P^K \to \frac{P_1}{2}$.

Observation 4 (i) and (ii) imply the NK system performs better than the K system and the benefit gained from using the NK policy tends to increase as the adoption rate of shelter 2 differs from shelter 1. From Observation 4 (iii), when the adoption rate of shelter 2 is very large, the killing probabilities of the K and the NK system converge to different values ($P^{NK} \to 0$; while $P^K \to \frac{P_1}{2}$). In the NK system, animals can be diverted if one shelter is full, so all the diverted animals arriving at the second shelter will not be rejected due to its large adoption rate. On the other hand, there is no coordination in the K system. When the second shelter's adoption rate is very high, it does not help reduce the killing probability of the first (fixed-adoption) shelter. The long-run killing probability of the described K system converges to $\frac{0+6P_1}{6+6} = \frac{P_1}{2} \approx 0.256$. Thus, the killing rate of the NK system can be effectively improved by enhancing the adoption rate of just one of the shelters in the system (which may be cheaper than enhancing the adoption rate of both shelters, for example).

Next, we consider some combinations of the K and NK policies since in practice it might not be possible to coordinate all shelters as a large NK system due to the distance between shelters (or the additional costs). We would like to explore coordination between different types of shelters for insights on which one performs better than the others. First, consider a system consisting of four shelters (C=30). Two shelters in the system are considered to be "Fast" shelters where the mean adoption time is equal to 8 and 16 days for good and bad-type animals, respectively. The other two shelters are "Slow" shelters with mean adoption time 12 for the good type and 24 for the bad type. The first system in consideration is called "F-F and S-S"; i.e., "Fast" shelter only coordinates with the "Fast" and the "Slow" only coordinates with the other "Slow" shelter. The second system is called "F-S and F-S" where the "Fast" shelters coordinate with the "Slow" ones. The left diagram of Figure 7 depicts the described systems, while the right diagram shows their killing probabilities at different arrival rates of the good-type (λ_1) and the bad-type animals (λ_2). The arrows indicate the pair of shelters that coordinate; while no arrows mean no coordinations between shelters.

Observation 5. *The "F-S and F-S" systems give lower killing probabilities than the "F-F and S-S" systems.*

Individually, the "Slow" shelter rejects a greater number of animals than the

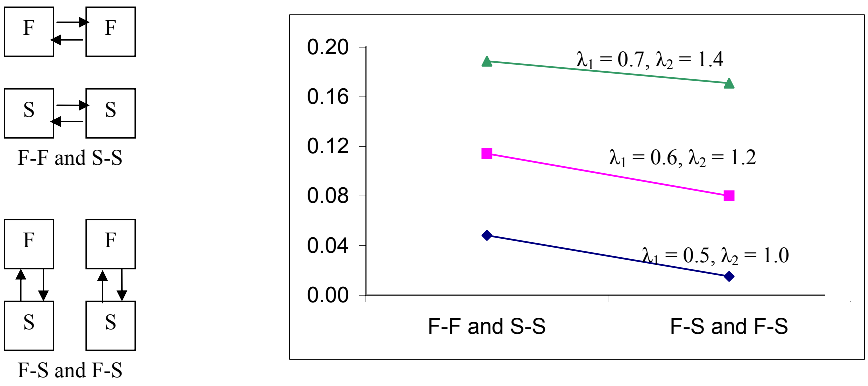

Figure 7: "F-F and S-S" and "F-S and F-S" systems (left) and their killing probabilities (right)

"Fast" shelter does. If a "Slow" shelter is paired with a "Fast" shelter, when the "Slow" shelter is full, animals are diverted to the faster shelter with higher adoption rates. In contrast, when a "Slow" shelter is diverting to another "Slow" shelter, they both may be full. Also, *if a set of shelters with different adoption rates are considering coordination but full coordination is too costly, then matching complementary shelters can be useful.*

Next, let us consider a system with three shelters where the inter-adoption times of the good-type animals are 10 days for "Fast", 12 days for "Medium" and 15 days for the "Slow" shelter, respectively. The inter-adoption times of the bad-type animals are assumed to be twice that of the good-type ones. Figure 8 presents a variety of policy combinations among the three shelters and their killing probabilities. Each pair of arrows in the figure represents coordination between shelters, and the letters beneath each picture denote the name of each system.

Observation 6.

i) For coordination between only one pair of shelters, the system that coordinates between the "Fast" shelter and the "Slow" shelters gives the lowest killing probability (i.e., "F-S" performs better than "M-S" and "F-M").

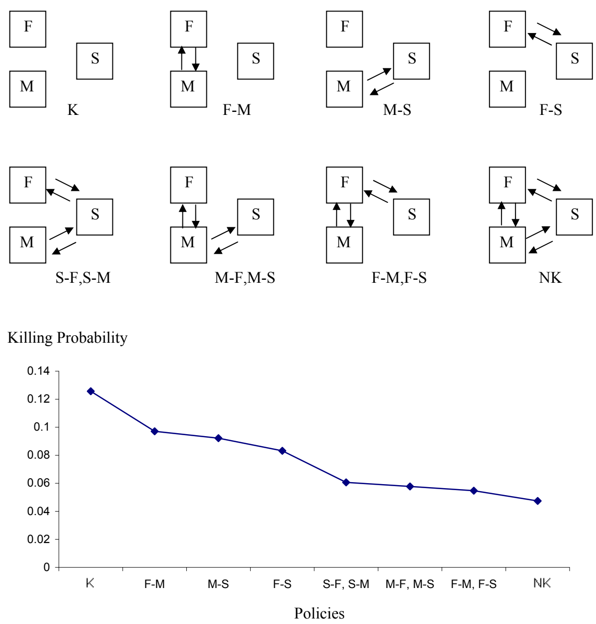

Figure 8: Two-way systems with "Fast", "Medium" and "Slow" shelters and their killing probabilities ($\lambda_1 = \lambda_2 = 0.5$/day)

ii) For coordination between two pairs of shelters, the system that coordinates between the "Fast" and "Slow" as well as "Fast" and "Medium" gives the lowest killing probability (i.e., "F-M,F-S" performs better than "M-F,M-S" and "S-F,S-M").

From Observation 6, if we have a variety of shelters in consideration and would like only one pair of coordination, selecting a pair with the fastest shelter and the slowest shelter is the most beneficial. In this case, we pair the shelter that needs help the most with the one that can help the most. For two pairs of coordinating shelters, the system connecting the fastest shelter to both of the others is likely to perform best. The reason is the "Fast" shelter is able to most effectively help reduce killing rates of both "Medium" and "Slow" shelters.

What we have explored in this section are "two way" coordinations; i.e., animals are diverted both forward and backward between shelters. In real life, some shelters may need to divert animals if they are full, but cannot receive animals from others due to their insufficient resources. Next, we consider what happens if "one way" coordinations are implemented. Figure 9 depicts systems with varieties of coordinations, where arrows indicate the direction of animals diversion. Parameters of "Fast", "Medium" and "Slow" shelters, e.g., arrival rates and adoption rates, are similar to those described in Figure 8.

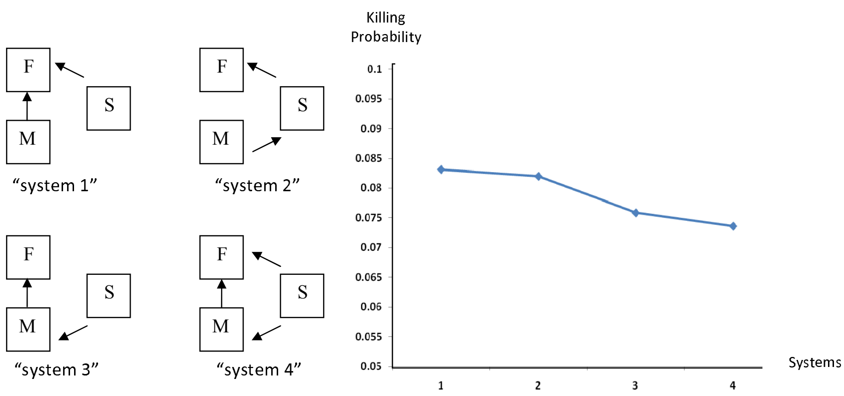

Figure 9: One-way systems with "Fast", "Medium" and "Slow" shelters when using "one-way" animal diversions

From Figure 9, "system 1" (both "Slow" and "Medium" shelters transfer animals to the"Fast" shelter) has the highest killing probability. A possible reason is when animals from all other shelters are diverted to the "Fast" shelter, they have to fight for spaces due to the diversion overload. Consequently, it hurts the system in the long run. The killing number can be much reduced by adding another diverting option from "Slow" to "Medium" shelter (as described in "system 4", which has the lowest killing probability). However, if there are only two pairs of shelters where the "one way" coordination is allowed, it is most effective to transfer animals from the "Slow" to "Medium" shelter, and from the "Medium" to "Fast" shelter ("system 3" in Figure 9). Therefore, *transferring animals from one shelter to the next faster one, e.g., "Small" to "Medium" and "Medium" to "Fast", can most help reducing the killing rate of the overall system (the killing probability of "system 3" is less than* *those of "system 1" and "system 2").* The examples we have explored in this section give insights on how to design an effective coordinating system in practice. If the best element is chosen for coordination, it can leand to much improvement in performance.

2.3.3.3 Systems of shelters whose the adoption rates are concave and increasing in the number of animals in the shelter

One of the assumptions we have used so far is that the shelter's adoption rate is proportional to the number of animals in the shelter, and we have found that NK performs better than the K system. In this section, we consider what happens if the shelter's adoption rate is not linear, specifically an concave increasing function of the number of animals in the shelter. In other words, the adoption rate still increases with the number of animals (due to a greater selection for people visiting) but the marginal benefit decreases. There are two questions of interest; 1) is the NK's killing rate still less than the K's killing rate?, and 2) if so, how does the function between the adoption rate and the number of animals affect the benefit of the NK policy over the K policy?

Let i be the number of animals in a shelter and $\nu(i, a, \mu)$ be the adoption rate when there are i animals in the shelter ($a \epsilon \{1, 2, 3, ...\}$). Define $\nu(i, a, \mu)$ as $\nu(i, a, \mu) = i^{\frac{a}{a+1}}\mu$, where μ is the adoption rate when there is only one animal in the shelter and "a" represents the sensitivity of the overall adoption rate over the number of animals. Note that $\nu(i, a, \mu)$ is increasing in a and is an increasing concave function of i. Further, since the function approaches i as a increases, this choice of function allows us to directly compare results to the case when the adoption rate is linear with the number of animals ($a \to \infty$).

Figure 10 (left) shows the shelter's adoption rate ($\nu(i, a, \mu)$) at different i, when a equals 1,3,5,7,9 and ∞, respectively. When $a \to \infty$, the adoption rate is closest to linear. Figure 10 (right) depicts the percentage decrease in the killing probability as the system of two shelters is operated by the NK policy instead of the K policy (when the total arrival rates to each shelter are 3, 3.5 and 4 per day, correspondingly).

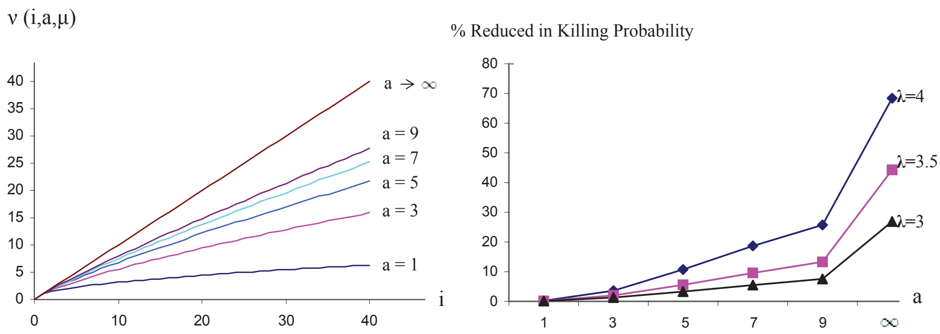

Figure 10: Shelter's adoption rate when μ equals 0.2 and a equals 1,3,5,7,9 and ∞ (left), and the percentage decreased in the killing rate as the two-shelter system is operated by the NK policy instead of the K policy (right)

Observation 7.

i) When the shelter's adoption rate is an increasing concave function of the number of animals in the shelter, the NK system's killing rate is less than the K system's killing rate.

ii) The percentage gained by using the NK policy in place of the K policy increases with a, i.e., as the adoption rate is closer to linear in the number of animals.

From Observation 7, the NK still performs better than the K system does with this non-linear assumption. As we discussed earlier, "a" is the indicator representing the sensitivity of the overall adoption rate over the number of animals in the shelter. When a is small, changing from K policy to NK policy may not offer as much benefit compared to when a is large. This insight suggests the shelter management to consider the sensitivity between the adoption rate and the number of animals when determining how much the system can be improved by the coordination policy.

2.3.3.4 Managerial Insights

In this section, we summarize managerial insights obtained from our computational experiments for coordination policy. For systems with identical mean adoption rates, we observe *the shelter cannot much reduce its killing rate by performing coordination when the existing killing rates are already high.* So, it might not be beneficial to focus on coordination in this case. We examine the percentage difference in adoption rates that would make the K system and NK system equivalent in their killing rates. For example, with a small acceptable killing probability, the adoption rate in the K system would need to be almost 10% faster. Therefore, *focusing on the adoption rate (e.g., using aggressive adoption campaigns) in the K system could achieve the same benefits as coordination.*

In addition, our result shows that when we increase the size of the NK system, the killing probability reduction is greater when the system is small than when the system is already large. If there is a cost of adding more shelters to coordinate, *the shelter manager should consider the trade-off between how much the system can improve and the cost of coordination (since having a very large NK system can be costly).* Instead of coordinating with all other shelters in the NK system, *coordinating only with shelters in neighborhoods, e.g., the "NK-neighbor" system, can also be an interesting and practical option.*

We found that *the NK system consisting of shelters with much different adoption rates can effectively reduce its killing rate by coordination policy.* For instance, if the manager can find a shelter with faster adoption rate to coordinate with a slower adoption rate, it can much reduce the overall killing rate of the system since animals diverted to the fast shelter have a higher chance finding adopters. In practice, when there are more than two shelters (e.g., "Fast", "Medium", and "Slow"), the manager may consider coordination among subsets of them. *If only one pair of coordination is allowed, for the highest improvement they should consider pairing the fastest shelter* *with the slowest one. For two pairs of coordination, the system connecting the fastest shelter with others gives the lowest killing rates.* Furthermore, we examine the impacts of "one way" coordination, where the animal diversion is allowed in one direction. We found *diverting animals from one shelter to the next faster shelter can most reduce the overall killing rate, compared to other options* (i.e., *"Slow" to "Medium" and "Medium" to "Fast"* is more effective than ' *'Slow" to "Fast" and "Medium" to "Fast"*, or *"Slow" to "Fast" and "Medium" to "Slow"*.)

Next, we relax the assumption that the shelter's adoption rate is proportional to the number of animals and consider what happens if the adoption rate is increasing concave with the number of animals in the shelter. The numerical experiments show coordination can still reduce the killing rate. However, its benefit is small when the adoption rate is less sensitive to the number of animals. So, *if the shelter management has observed that the total adoption is not much impacted by the number of animals in the shelter, coordination may not be a good option to improve the system.*

In this section we have explored effects of coordination among shelters. In the next section, we study how No-Kill shelters can reduce the killing rate by adoption and neutering campaigns.

2.4 Adoption and Neutering Policies

In addition to shelter coordination, No-Kill shelters can reduce their killing rates by encouraging more adoptions and reducing animal arrivals. These two practices are often implemented in No-Kill shelters [10]. We consider adoption campaigns and neutering campaigns in Section 2.4.1 and 2.4.2, respectively, and the comparison of campaigns is presented in Section 2.4.3.

2.4.1 Adoption Campaign

An adoption campaign is often promoted to increase adoption rates. In this campaign, the shelters use, for instance, temporary storefronts, mobile units, partnerships with retail operations via the internet and etc. [148]. The campaigns usually raise not only the number of animals who are placed in new homes, but also the return-to-owner rate for lost pets. In this section we investigate the impact of the number of adoption campaigns on the shelters' overall cost and benefit. An important decision for shelters' management is how many adoption campaigns should be used per year [97]. To consider this, define $\mu_1(n_a)$ and $\mu_2(n_a) = k\mu_1(n_a), 0 < k < 1$, as the adoption rates of the good-type and the bad-type animals, respectively, where n_a is the number of adoption campaigns that a No-Kill shelter attempts in one year.

Like the previous section, we consider λ_1 and λ_2 as the yearly arrival rates of good-type and bad-type animals, accordingly. Let v_k be the cost of killing an animal, which can be interpreted as the disadvantageous impact of euthanizing each animal, f_a be a fixed cost and v_a be a variable cost of using adoption campaigns. The shelter's total cost when implementing an adoption policy with n_a campaigns can be defined as

$$TC_a^{NK}(n_a) = v_k(\lambda_1 + \lambda_2)P_C(\mu_1(n_a)) + (f_a + v_a n_a), \tag{1}$$

where the shelter's killing probability, $P_C(\mu_1(n_a)) = \frac{\frac{(\frac{\tilde{\lambda}}{\mu_1(n_a)})^C}{C!}}{\sum_{i=0}^{C}\frac{(\frac{\tilde{\lambda}}{\mu_1(n_a)})^i}{i!}}$ and $\tilde{\lambda} = \frac{k\lambda_1+\lambda_2}{k}$.

Note that we consider only the costs directly impacted by adoption campaigns. The first part of equation (1) is the killing cost, which can be reduced by increasing adoptions, and the second part represents fixed and variable costs of implementing the campaigns. The main notation we employ in this section is provided in Table 4.

Lemma 2. *If the adoption rate, $\mu_1(n_a)$, is concave in the number of adoption campaign, n_a, then the shelter's total cost, $TC_a^{NK}(n_a)$, is a convex function with respect to n_a.*

Since there is a limitation on the number of households near each shelter, clearly the impact of each adoption campaign on the adoption rate is non-increasing concave with the number of campaigns a shelter has attempted ($\mu_1(n_a)$ is concave in n_a). From

Table 4: Notation for Section 2.4.1

n_a	The number of adoption campaigns
$\mu_1(n_a)$	Adoption rate of the good-type animals when the shelter promotes n_a adoption campaigns
v_k	Cost or negative impact per an animal killed in the shelter
f_a	Fixed cost of implementing the adoption policy
v_a	Variable cost of implementing each adoption campaign
$TC_a^{NK}(n_a)$	Total cost of a No-Kill shelter implementing adoption campaign policy
$\overline{n}_a$	The maximum number of adoption campaigns due to the budget constraint

Lemma 2, the convexity of the shelter's total cost function allows us to determine the optimal number of adoption campaigns as discussed in the following Theorem, where $\overline{n}_a$ is the maximum number of campaigns a shelter can use (e.g., due to budget limitation).

Theorem 3. *The optimal number of campaigns in the interval* $[0, \overline{n}_a]$ *for all* $n_a \in [0, \overline{n}_a]$ *is given by :*

i) $n_a^* = 0$ *when* $v_k[\lambda_1 + \lambda_2]P_C(\mu_1(0))\Big[\frac{C\mu_1(0)}{\tilde{\lambda}} - 1 + P_C(\mu_1(0))\Big]\frac{\tilde{\lambda}}{[\mu_1(0)]^2}\frac{\partial \mu_1(0)}{\partial n_a} < v_a$*,*

ii) $n_a^* = \overline{n}_a$ *when* $v_k[\lambda_1 + \lambda_2]P_C(\mu_1(\overline{n}_a))\Big[\frac{C\mu_1(\overline{n}_a)}{\tilde{\lambda}} - 1 + P_C(\mu_1(\overline{n}_a))\Big]\frac{\tilde{\lambda}}{[\mu_1(\overline{n}_a)]^2}\frac{\partial \mu_1(\overline{n}_a)}{\partial n_a} > v_a$*,*

iii) the solution of the following equation, otherwise:

$$v_k[\lambda_1 + \lambda_2]P_C(\mu_1(n_a))\Big[\frac{C\mu_1(n_a)}{\tilde{\lambda}} - 1 + P_C(\mu_1(n_a))\Big]\frac{\tilde{\lambda}}{[\mu_1(n_a)]^2}\frac{\partial \mu_1(n_a)}{\partial n_a} = v_a. \quad (2)$$

Note that case (i) implies the shelter should not promote any adoption campaign since the benefit gained from the reduced killing rate is always lower than the cost of campaigning. However, this condition rarely holds for shelters in real-life situation because the killing rate (without any adoption campaign) times killing cost is usually higher than the cost of employing an adoption campaign [97]. In case (ii), the total cost is still decreasing in n_a when $n_a = \overline{n}_a$. The shelter can benefit more with an additional number of campaigns but they already reach the maximum $\overline{n}_a$. While in case (iii) we have the total cost first decreases and then increases in the interval $[0, \overline{n}_a]$ and n_a^* is the solution of equation (2).

To illustrate impacts of adoption campaign on the total cost, consider the following example of $\mu_1(n_a)$ function:

$$\mu_1(n_a) = w_a + z_a(1 - e^{-l_a n_a}), \qquad w_a, z_a, l_a > 0.$$

It is important to note that when the shelter does not implement any adoption campaign, the starting adoption rate of this function is $\mu_1(0) = w_a$. Moreover, the increased adoption rate does not exceed z_a ($\lim_{n_a \to \infty} \mu_1(n_a) \to w_a + z_a$) no matter how many campaigns implemented. Also, it can be easily verified that $\frac{\partial \mu_1(n_a)}{\partial n_a} > 0$ and $\frac{\partial^2 \mu_1(n_a)}{\partial n_a^2} < 0$.

Let us consider the example of a shelter with capacity 80 where the adoption rate is 15 without any campaigns and it cannot exceed 25 no matter how many campaigns attempted ($w_a = 15, z_a = 15$ and $l_a = 0.2$). Since the average cost of basic food, supplies, medical care and training for a dog or cat is 400 to 700 dollars annually [110], we approximate the killing cost to be 400 in this example. Let the campaign's fixed and variable cost be 2000 and 1000, respectively. Figure 11 shows the shelter's costs as a function of the number of adoption campaigns at different yearly arrival rates of bad-type animals (the rate of good-type animals is assumed to be 70% of the bad types' in this example). The shelters with greater arrival rates have higher minimum costs and need a larger number of campaigns to achieve them. As Figure 11 implies the arrival rate plays a role in how much the shelter can lower its cost, in the next section we consider another No-Kill solution, namely neutering campaigns.

2.4.2 Neutering Campaign

Although adoption is important, it alone cannot effectively solve animal shelter's problems. The number of animals coming into the shelters needs to be reduced. In

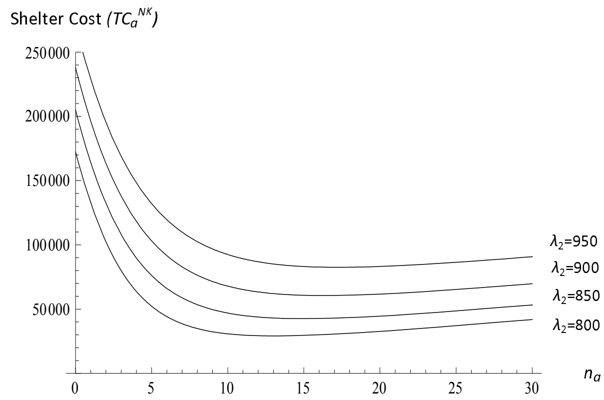

Figure 11: The total cost of shelters implementing adoption policy, at different number of adoption campaigns, when $\lambda_2 = 800, 850, 900$ and 950, respectively

typical service systems (e.g., call centers) reducing the arrival rates is not a common practice since it means fewer customers (and usually less revenue). However, for non-profit organizations like animal shelters, having higher arrival rates can cause higher number of euthanized animals, so decreasing arrival rates is desirable.

Spaying or neutering is one of the most effective ways to solve pet overpopulation and eventually reduce the shelters' killing rate [132]. Unfortunately, the current statistics shows that only 10% of the animals received by shelters have been spayed or neutered [110]. Therefore, No-Kill shelters widely encourage pet owners to neuter their animals via neutering campaigns, by offering reduced rates or free services for people bringing animals to be neutered. Although there are benefits of future reduced incoming animals to shelter when an animal is spayed, there exist costs of neutering processes. For the shelter's effective decision making, it is important to examine the impact of the number of neutered animals on the shelter's total cost. Let $h\lambda(n_s)$ and $\lambda(n_s)$ be the yearly arrival rates of good-type and bad-type animals, respectively, where $0 < h < 1$ and n_s is the number of animals spayed each year. It is logical to assume the arrival rate of bad-type animals is greater than the good type since they have a higher chance of being left in the shelters. Also, for neutering campaigns, shelters usually neuter either types depending on which one is brought to them, so we assume each animal neutered has impact on both types' arrival rates. The shelter's cost when using a neutering (or spaying) campaign is defined as $TC_s^{NK}(n_s)$ and

$$TC_s^{NK}(n_s) = v_k[h\lambda(n_s) + \lambda(n_s)]P_C(\lambda(n_s)) + (f_s + v_s n_s), \tag{3}$$

where $P_C(\lambda(n_s)) = \frac{\frac{(\frac{h\lambda(n_s)}{\mu_1}+\frac{\lambda(n_s)}{k\mu_1})^C}{C!}}{\sum_{i=0}^{C}\frac{(\frac{h\lambda(n_s)}{\mu_1}+\frac{\lambda(n_s)}{k\mu_1})^i}{i!}}$; f_s and v_s are the fixed and variable costs of neutering policy, respectively (see Table 2.4.2 for summary of notation).

Table 5: Notation for Section 2.4.2

n_s	The number of spayed animals due to neutering campaigns
$\lambda(n_s)$	Arrival rate of the bad-type animals when n_s animals neutered in total.
$h\lambda(n_s)$	Arrival rate of the good-type animals when n_s animals neutered in total.
f_s	Fixed cost of implementing neutering campaigns
v_s	Variable cost of neutering each animal
$TC_s^{NK}(n_s)$	Total cost of a No-Kill shelter implementing neutering campaigns

Lemma 3. *If the arrival rate, $\lambda(n_s)$, is non-increasing convex in the number of spayed animals, n_s, then the shelter cost, $TC_s^{NK}(n_s)$, is a convex function with respect to n_s.*

Lemma 3 gives a sufficient condition for the shelter cost to be convex in the number of spayed animals. For some service systems, it is reasonable to consider the reduced arrivals as a convex function. However, for the animal shelter system, it may be necessary to consider population dynamics. For example, in practice $\lambda(n_s)$ may not always be convex in n_s. The motivation of this assumption comes from the well-known population growth logistic function. In other words, it is likely that when there are very few number of spayed animals at the beginning, the arrival rates into the shelter will not be reduced very much. Once there are greater number of spayed animals, the arrival rates can be decreased with an increasing rate until a limit when there are too many neutered animals in which the arrival rate is already low and cannot be much more reduced. For example, $\lambda(n_s)$ can be non-increasing first concave and then convex function of n_s. In addition, animal experts state that spaying and neutering as much as 70% of the animals in a pet overpopulation area is necessary for significant results in reducing strays [132]. So a sufficient number of animals must be neutered for an effective population reduction. In our work, we focus on effects of the number of neutered animals rather than effects of time, which is often seen in the analysis of a general growth function.

To determine the impact of n_s on the shelter's total cost, we consider the following example of $\lambda(n_s)$ function:

$$\lambda(n_s) = \frac{w_s(w_s + z_s)}{w_s + z_s e^{l_s n_s}}. \tag{4}$$

Although $\lambda(n_s)$ can take other forms, the arrival rate function in (4) is logical for the following reasons. Firstly, when there are no animals neutered, the arrival rate starts at w_s (or $\lambda(0) = w_s$). Secondly, when all animals are neutered, the arriving rate approaches zero (i.e., $\lim_{n_s \to \infty} \lambda(n_s) \to 0$). Thirdly, it is easy to see that $\lambda(n_s)$ is strictly decreasing with n_s ($\frac{\partial \lambda(n_s)}{\partial n_s} < 0$), concave when $0 \leq n_s < \frac{\log \frac{w_s}{z_s}}{l_s}$ and convex when $n_s \geq \frac{\log \frac{w_s}{z_s}}{l_s}$. Parameters w_s, z_s and l_s can be chosen to best reflect the effects of the number of animals neutered on the arrival rates of each shelter. To analyze impacts of neutering campaigns, we address key structures of the shelter's cost function in the next property.

Property 4. *If the arrival rate function* $\lambda(n_s) = \frac{w_s(w_s+z_s)}{w_s+z_s e^{l_s n_s}}$*, we have:*

i) the shelter cost, $TC_s^{NK}(n_s)$*, is convex in* n_s *when* $n_s \geq \frac{\log \frac{w_s}{z_s}}{l_s}$*,*

ii) there exists at most one $n_s^* < \frac{\log \frac{w_s}{z_s}}{l_s}$ *such that the shelter cost,* $TC_s^{NK}(n_s)$*, is first concave when* $n_s < n_s^*$ *and then convex when* $n_s \geq n_s^*$*.*

Figure 12 shows the shelter's cost as a function of the number of animals neutered when $\mu_1 = 20$, 23, 26 and 29, accordingly[1]. At small n_s, the total costs are concave

[1] Parameters for Figure 12 are w_s=1200, l_s=0.01, z_s=20, f_s=300 and v_s=60.

decreasing and then change to be convex increasing as n_s increases. Furthermore, the total costs increase as n_s goes to ∞. So when only very few animals are spayed, the total cost cannot be reduced much, but the effect become more significant once the number increases. However, when too many number of animals are spayed, the neutering cost offsets the benefit gained in the reduced killing cost, so the total cost increases as more animals are spayed. From Figure 12, the effect of the neutering campaign is greater for shelters with a small adoption rate. Nevertheless, they need to have a large number of spayed animals to reach the minimum cost.

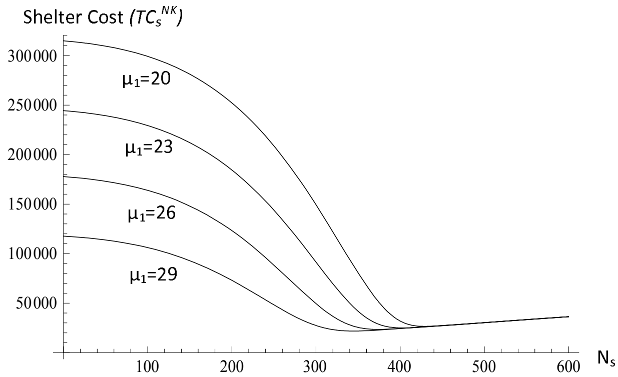

Figure 12: The cost of shelters with neutering policy as functions of the number of animal neutered when μ_1 = 20, 23, 26 and 29, respectively

2.4.3 Campaigns Comparison

One of the differences when comparing adoptions to neutering policy is: adoption campaigns can reduce the shelter cost quite significantly with a small number of campaigns (since the cost is convex in n_a), while often, neutering improves the cost slowly for a small investment (since the cost might be concave with n_s). However, both policies are similar when high effort is used since the shelter cost increases with the effort ($\lim_{n_a \to \infty} TC_a^{NK}(n_a) = \lim_{n_s \to \infty} TC_s^{NK}(n_s) \to \infty$). This is the case because the benefit gained cannot be offset by the campaign cost at the extreme. Examining the two policies raises another interesting question: if a No-Kill shelter has a fixed amount of budget, under which scenario is one policy more effective than the other?

Consider a shelter that has a fixed S dollars budget and wants to spend on either neutering or adoption or both policies at the same time. Figure 13 displays the shelter total cost for dollars spent on the following: (1) adoption only, (2) 67% adoption and 33% neutering, (3) 33%adoption and 67%neutering, and (4)neutering only[2].

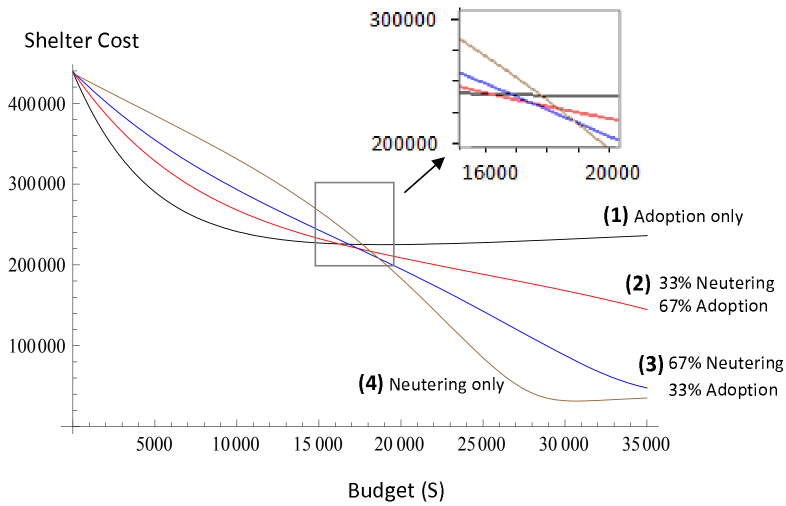

Figure 13: The shelter's overall cost for dollars spent on: (1) adoption only, (2) 33%neutering and 67%adoption, (3) 67%neutering and 33%adoption, and (4) neutering only

From Figure 13, the first policy (adoption only) is able to more effectively reduce the overall cost when the total budget is less than approximately 16,000 dollars (as the total cost of policy (1) is the lowest line). When the budget is between 16,000 and 19,000 dollars, using both adoptions and neutering campaigns (policy (2) and (3)) can bring shelter cost to be lower than using only one campaign (policy (1) and (4)). However, if the shelter is able to spend more than 19,000 dollars, the overall cost can be much reduced with a neutering policy (4). Note that the minimum cost occurs when shelter spends approximately 29,000 dollars with policy (4), in which

[2]Parameters for Figure 13 are w_a=15, z_a=10, l_a=0.2, f_a=2000, v_a=1000, w_s=1200, l_s=0.01, z_s=20, f_s=300 and v_s=60.

the shelter implements only the neutering campaigns. Although it is not shown, the overall cost of all policies increases when the investment is high, since the benefit gained cannot offset the campaign costs.

Our analysis has shown that one policy can be more beneficial than others under different situations, e.g., depending on the shelters' the offered load (r) and the budget constraint. Figure 14 summarizes our observations on the policies that the shelter management should emphasize, where the horizontal axis represents budget limitation and the vertical axis represents the offered load (i.e., the ratio between arrival and adoption rates). Although the killing rate is small at low r, neutering policy can be useful because it prevents future animal overpopulation problems. Shelter coordinations are beneficial for medium r, but not for small or high r, as shown in Property 3 that the successful diversion rate approaches zero in those two cases. For high r, adoption and neutering policies are necessary to increase shelter's turnover rates and reduce incoming rates, respectively. As we have seen in Figure 13, adoption campaign is favorable when the budget is low; while neutering policy can be a good option when the budget is high enough to spay sufficiently large number of animals. Moreover, combination of adoption and neutering campaigns is useful for a moderate campaign budget. If the shelter management emphasizes on the right policy (depending on their situations), the shelter's problem can be most effectively solved.

2.5 Conclusions

Approximately five million cats and dogs were killed in animal shelters last year in the United States. With such a large number, shelters are facing a challenging problem of how to decrease the number of animals euthanized. Some shelters have moved to a No-Kill policy, where their animals are not killed for space reasons. Instead, they try to solve the problem by using several methods. This chapter explored different kinds of No-Kill main policies: coordination with other shelters, adoption and neutering,

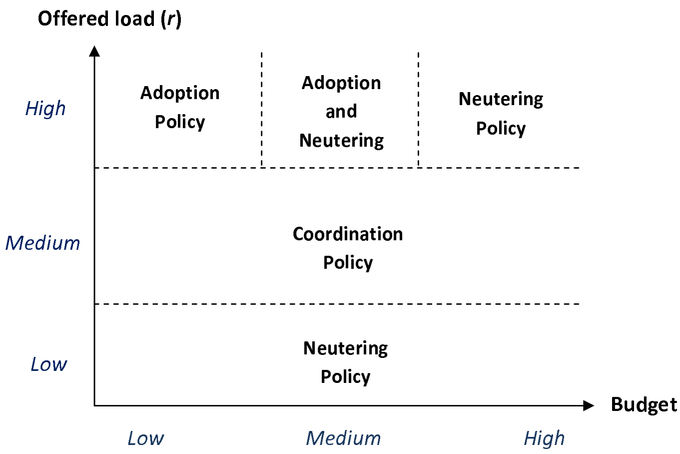

Figure 14: Summary of beneficial policies under different scenarios

with the goal of providing insights on how each method helps reduce the killing rates (or reduce the job rejection rate of other service systems).

For the coordination policy, if one shelter is full, animals can be diverted to the other shelters in the NK system. We addressed the theoretical aspects of coordination and used numerical analysis to obtain managerial insights. We proved that the coordinating NK system with m shelters having the same adoption rates is equivalent to the $M/M/mC/mC$ system. This directly allowed us to show the NK system has a lower killing rate than the uncoordinated K system. The systems may have two types of animals, where the animals may arrive or be adopted at different rates. Analytically, we showed changing the arrival rates for the bad type had more impact on the overall system than changing the arrival rate of the good type (independently of their relative frequency of arrivals). We also showed that the killing rate of the shelter decreased not only in individual adoption rates, but also in the ratio of the adoption rate of the bad-type to the good-type.

When the adoption rates of each shelter are not identical, we found the NK system was no longer equivalent to the $M/M/mC/mC$ system. This makes theoretical results more difficult to show, but we were able to use numerical analysis to generate further insights. We found that *adding more coordinated shelters helps reduce the overall killing probability but the marginal improvement is decreasing.* Thus, if there is a fixed coordination cost per shelter added, *it might not be optimal to have a large NK system.* We provided several insights on how to design different kinds of coordinating systems to achieve the highest improvement, including coordinating with neighbors or between particular pairs of shelters.

Our results show that coordination cannot much help reduce the killing rate when the existing rejecting probability is high. Adoption and neutering policies can be useful in that case. We showed when a shelter uses adoption campaigns, *if adoptions are concave and increasing with the number of the campaigns, the shelter's total cost is a convex function with the number of campaigns and there exists an optimal number to achieve the minimum cost.* For the neutering campaign we found *when animals increase according to a type of growth function, the shelter's total cost is first concave and then convex in the number of neutered animals.* Comparing these two policies, computational experiments suggested that an adoption campaign could be more attractive when the budget is limited, while neutering policy could be a good option when the shelters have enough money to spay sufficiently large number of animals. Since it may not be possible for shelters to apply all policies at once, we discussed the situation when each policy is the most effective tool. Although this work is motivated by the animal shelters' policies, the context can be applied to other industries with customers and servers, e.g., hospital operations, ambulances and internet servers. Finally, we hope our results will be helpful as we strongly believe that killing is not the only option to solve the shelters' problem.

CHAPTER III

SCALING THE HOUSE : OPTIMAL SEATING ZONES FOR ENTERTAINMENT VENUES WHEN LOCATION OF SEATS AFFECTS DEMAND

3.1 Introduction

Revenue management (RM) has been a significant strategy to maximize revenue for industries with limited resources. RM research was pioneered in the airline industry and has been widely practiced in other industries such as hotels, cruise lines and rental car agencies (e.g., [7], [25], [55], [82], [91], [94]). Sports and entertainment (S&E) is another of industry with great potential for applying the revenue management practices in business decisions. However, much remains to be done in developing new methods and approaches to apply to the revenue management in the S&E industry [147].

There are several important decisions that the S&E venue manager needs to make: e.g., how to categorize seats into zones for different prices, what price they should charge for each seating zone, and which performances and how many should be included in a ticket bundle. In this chapter, we focus on developing methods to systematically "scale the house" as it is called in the industry, which is sectioning seats into zones for different prices. Our work arose from discussions with a firm who provides analysis of revenue management for many large performance venues. They point out that "scaling the house" is one of most important decisions the venue managers need to make. Although several studies in the literature indicate price differentiation can increase performance venues' revenue (e.g., [67], [89] and [120]), determination of the seating categories has not been deeply explored [34].

Tickets sold in the S&E industry differ from the standard spot market since tickets of the same show or game are not homogeneous items. Each seat in the venue offers a different experience as it is located in a different location, so customers perceive them as products with different qualities. However, it is not practical to price each individual seat with a unique price. General guides state that seats closest to the stage are usually priced at a premium, and the price falls as they get closer to the back of the venue [5]. While most of the previous studies related to demand estimation for performing arts focus on finding the impact of related variables, such as price, on the aggregated demand (e.g., [33], [43] and [47]), we incorporate impacts of seat locations and then integrate those impacts to identify an optimal zoning decision for the venue.

In this study, we first examine transactional level data obtained from major arts organizations to identify how ticket demand varies by location of seats. In addition to price, we find that demand decreases with distance from the stage and distance from the seating row's center, where the highest demand occurs in the center-front areas. Given the observed demand characteristics, we consider how the venue manager should divide seats into zones to maximize the expected revenue. We focus on revenue for this industry because the cost of selling additional tickets is very small compared to the performance's fixed set-up cost.

We consider a static scaling the house decision because most art organizations publish their seating sections early in the selling horizon and prefer not to dynamically change them. The reasons are as follow: 1) most advertising brochures contain the seating charts indicating how seats are categorized, so changing seating zones after they have been announced is not practical; 2) some venues may have the zone name printed on the tickets or may use different ticket colors to differentiate zones. For these reasons, finalizing the zoning decision at the beginning of the selling season is preferred.

We develop a "Scaling the House" model to segment the venue across two dimensions (e.g., by row and column cuts), incorporating demand differences from front or center. For venues whose ticket demand is not very sensitive to the distance from the center, we present an alternative one-dimensional optimal zoning decision where sections are defined by seating row cuts, which provides an easy-to-implement optimal decision. In both models, we analytically identify the optimal solutions and provide comparative statics of results for managerial insights. We perform robustness analysis to examine how model results are sensitive to estimation errors under different scenarios. We also conduct several computational experiments to identify the situations when it is the most beneficial to apply the two-dimensional model rather than the one-dimensional zoning decisions. It is worthwhile to note that our models and results also provide some implications to other industries whose demand is sensitive to spacial locations such as airline seats, etc.

This chapter is organized as follows. In the next section, we provide a review of the related literature. In Section 3.3, we present empirical results on how seating locations affect demand within venues. We describe the resulting demand function based on location in the venue and present the two-dimensional zoning model where zones are divided from the front to the back as well as from the center to the left/right in Section 3.4.1. Then, in Section 3.4.2, we present an alternative zoning model where zones are defined in only one dimension. Robustness analysis and comparisons between the different zoning models are considered in Section 3.5 and 3.6, respectively. We conclude by summarizing important results and providing managerial insights in Section 4.6. The proofs of all results are provided in the Appendix.

3.2 Literature Review

A stream of literature, especially from economics, related to the sport and entertainment (S&E) industries focuses on ticket demand estimation. In an early paper, Moore [106] developed an econometric model of performance theater attendance. He focused on finding the effects of ticket price, income, and the number of shows on ticket demands. Felton [47] estimated a demand function for opera tickets and used a "popularity rating" variable as a measure of quality. Ekelund and Ritenour [43] incorporated an opportunity cost variable (mean hourly wage across cities) in their model and found a significant negative impact on the aggregate symphony concert demand. Corning and Levy [33] considered demand for performances in a group of live theaters and characterized the seasonality of demand for each theater. Tseng [140] empirically explored how consumers respond to event manager's scheduling decisions. He found that when performances are scheduled closely in distance or time can sometimes have either positive or negative effects on ticket sales. Other empirical works in ticket demand include [19], [58], [74], and [138]. All of the studies above used aggregated (time series or cross-sectional) data to find the effects of relevant factors (such as income and price) on the aggregate ticket demand. In contrast, our work quantifies demand variation within the same venue based on locations, such as the distance of seats from the stage and the distance from the center of each seating row. We observe how customers perceive seats at different locations in the venue, which cannot be obtained from the aggregated demand data.

In addition to ticket demand estimation, studies related to the S&E business include ticket bundling and pricing of complementary goods. Regarding a ticket bundling problem, Drake et al. [41] developed an optimal switching model from selling ticket bundles to selling a single ticket, and identified the conditions when only bundles, only individual tickets, or both should be offered. Duran and Swann [147] extended the static model to a switching model that considered the optimal dynamic time to switch based on inventory. Their computational experiments suggested that the profit improvement could be 1 to 2 percent over the optimal static switching and the dynamic switching could also reduce profit variability. Studies on pricing complementary goods are presented in [100] and [134]. For example, Marburger [100] found that when the performance price setter gains a share of revenues from complementary goods, overall profits is maximized when the performance ticket prices lie in the inelastic section of demand. None of the papers above considered variable pricing of seats, while our work incorporates seating categorizations in the venue with the goal to increase ticket revenue via effective "Scale the house" decisions.

Our work also relates to price discrimination literature. A body of evidence indicates that price discrimination plays a significant role in the S&E industries. Huntington [67] examined a sample of thirty-three theaters to compare the revenue between a single price policy and a multiple price policy. From regression results, he suggested that theaters charging a uniform price could increase their revenue by approximately 24 percent by scaling the house. Leslie [88] analyzed the existence of price discrimination for a particular Broadway show. He found that there are, on average, 8.7 different ticket categories sold for each show. In addition, Leslie [89] conducted various experiments to examine the implication of different pricing mechanisms and showed price discrimination can increase profitability by approximately 5 percent, compared to uniform pricing across venue. The author assumed the ticket quantity of each price section is fixed, so he did not consider the "Scaling the house" decision in his model. Advantages of multiple seating categories are also pointed out in Courty [36], where he empirically found that offering various prices could cause revenue to be about 5 percent higher than offering a uniform price. Other studies focused on the variations on ticket price discrimination can be found in [42], [56] and [95]. Although many papers have found significant advantages of ticket price segmentation, none of the authors above considered methods to optimally segment seats into zones to maximize revenue.

Although different seats in the S&E venue provide different views of the stage, each seat is seldom priced as a distinct good. Instead, the venue manager usually groups seats in sections or categories and prices them equally within the same section ("Scaling the house"). Rosen and Rosenfield [123] considered an economic model with two seat categories and determined the number of tickets for each category . Unlike our work, they did not include the impact of spatial characteristics (e.g., locations of seats) when determining the number of tickets. Since we introduce an analytical method of scaling seats into different sections, our work contributes to the current literature in finding optimal seating zones based on their location in the venue.

In summary, unlike most previous empirical studies that employed aggregated demand data, we explore the detailed transactional data to identify how demand varies with location. We quantify the effect of the distance from the stage and the distance from the center on ticket demand and develop a two-dimensional "Scaling the House" model to find the optimal points to switch from a premium price to a lower price (from front to back, and from center to the left/right of the venue). When demand is not very sensitive to the distance from the center, we alternatively consider a one-dimensional model focusing on seats sectioning only from front to back. We characterize the optimal venue segmentation accounting for location effects nd analyze insights for managements. Our results shed light on "scaling the house" strategies that had not been spotlighted by the existing literature.

3.3 Effect of Seating Location on Ticket Demand

In this section, we examine how seating locations affect demand within a venue. We use detailed transactions of classical concert ticket sales ranging from year 2004 to 2009, obtained from a major performance venue in the U.S. in this study and we have 318 shows across 5 seasons. The detailed transactional data that we study contains information such as ticket seating locations, ticket prices and when the transactions occurred. The venue capacity is 1,762 and the average percentage seat sold is 74.4 percent. In addition to this data set, we also examine ticket demand at two other large venues and find similar results. The seating chart of this venue is presented in Figure 15.

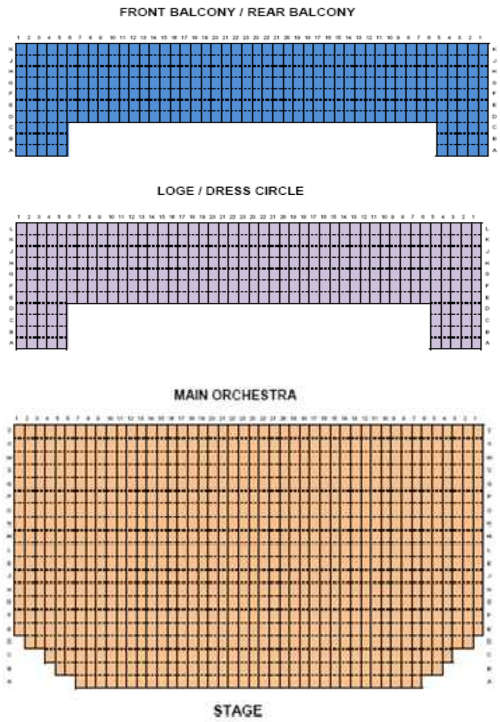

Figure 15: Seating chart of the venue

To measure the average demand of seats by location, we use the average percentage sold of each seat across each season as the ticket demand indicator. Note that although the analysis could also be done for shows in the season of a given type (e.g., symphony or dance) if desired. Zoning decisions in the S&E industry are usually made before the performance (or game) season begins and seating sections do not vary for each performance of the same type (i.e., classical, pop, etc.). Therefore, the average percentage of seats sold during the whole season can represent the overall popularity of each location in the venue. Since there are many artists or games that were performed at the same venue each year, the average percentage sold represents the seat popularity in general. In addition, since none of the performances in the data are totally sold out, the percentage of seats sold generally represents ticket demand in this venue. Alternative measures of demand can be used for venues that sell out, such as how quick each seat is sold.

As an example of the ticket demand, Figure 16 presents the percentage of tickets sold in the two latest seasons for different seating rows. From Figure 16, "07-08" and "08-09" indicate 2007-2008 and 2008-2009 seasons, respectively, where each season starts in September and ends in June or July of the following year. Note that as the number of rows (at the horizontal axis) increases, seats are further from the stage. We can see in Figure 16 that the percentage sold significantly decreases with the distance from the front for both seasons, suggesting the distance negatively affects ticket demand.

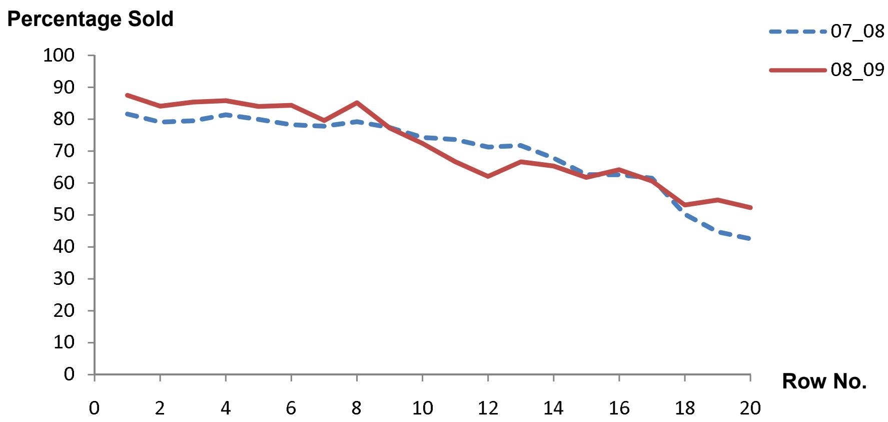

Figure 16: The percentage of seats sold, located at different rows in the entertainment venue, where the front of the venue is near row 0

Figure 17 shows the percentage of tickets sold at different seating columns, from the left to the right of the stage. We can observe the highest percentage sold occurs in the middle part of the venue and demand is generally lower as seats are further away from the middle or center of the seating rows. Moreover, compared to nearby locations, the percentage sold is also higher for seats next to the aisle, which are the very first and very last two columns.

To identify the overall effect of locations on the demand from the data, we consider

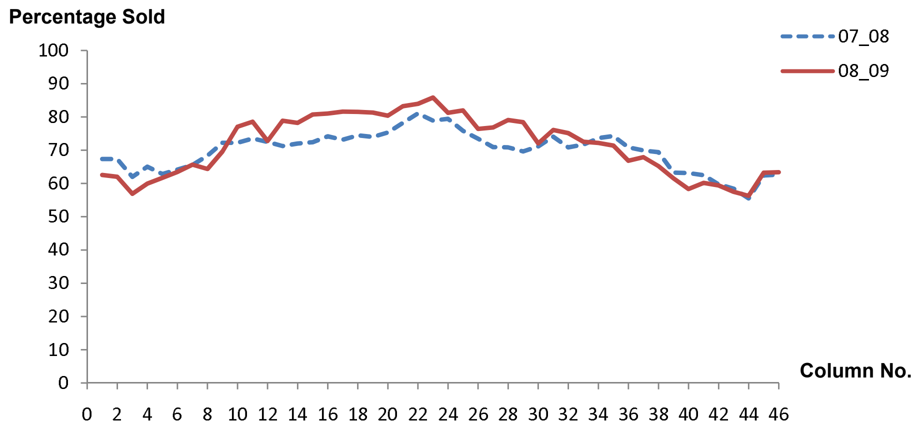

Figure 17: The percentage of seats sold, located at different columns in the entertainment venue, where the left of the stage is column 0)

the following estimation model,

$$Percent_{sold,l} = \theta + \beta_p P_l + \beta_d D_{front,l} + \beta_c D_{center,l} + \beta_a Aisle_l + \sum_{k=2}^{5} \gamma_k S_{k,l} + \epsilon, \quad (5)$$

where ϵ is an error term. The dependent variable $Percent_{sold,l}$ is defined as the average percentage of seat l that is sold during each season. The independent variable P_l denotes the average price that the customers actually paid for seat l, including discounts that may have been received. Variables $D_{front,l}$ and $D_{center,l}$ denote the distance from the front and the distance from the center of seating rows of seat l, respectively; and $Aisle_l$ is the variable indicating if that seat is next to aisles. The dummy variable, $S_{k,l}$, indicates which season the data is from (where the first season, 2004-2005, is used as the starting point), and $k = 2, \ldots, 5$ represent 2005-2006,..., 2008-2009 seasons, respectively) The parameters θ, β_p, β_d, β_c, β_a and $\gamma_k(k = 2, \ldots, 5)$ are to be estimated. To summarize the data characteristics, we present the descriptive statistics of each variable in Table 6.

The model fit is presented in Table 7 to describe effects of price, location of seats, etc., where the R-squared of the estimation model is 0.615. We find that when the price of seats increases by 4 dollars, the percentage sold decreases by approximately one unit (e.g., from 80% to 79%). Both distance from the front and distance from the

Table 6: Descriptive statistics

Variable	N	Mean	SD	Min	Max
$Percent_{sold,l}$	2209	74.41	22.78	6.35	99.25
P_l	2209	29.75	7.13	9.82	46.65
$D_{front,l}$	2209	9.17	5.61	0.00	19.00
$D_{center,l}$	2209	10.78	6.61	0.00	22.00
$Aisle_l$	2209	0.08	0.28	0.00	1.00
$S_{2,l}$	2209	0.20	0.40	0.00	1.00
$S_{3,l}$	2209	0.20	0.40	0.00	1.00
$S_{4,l}$	2209	0.20	0.40	0.00	1.00
$S_{5,l}$	2209	0.20	0.40	0.00	1.00

center have negative impacts on the percentage sold, but the impact of the distance from the front is approximately twice as much as that of the distance from the middle (-1.94 versus -0.96). The results imply that demand is relatively more sensitive to the distance from the stage in this venue; i.e., paying the same price, customers may prefer moving one seat from the middle section to moving further to the back of the venue. In addition, we show that aisle seats have positive impacts on sales, although the majority of seats are not by the aisles. The percentage sold is significantly higher for the last three seasons (2006-2007, 2007-2008 and 2008-2009) than the first two seasons (2004-2005 and 2005-2006), where the highest sales occur in the last 2008-2009 season. This analysis is relatively easy to repeat for the particular venues where the type of event may drive more or less sensitivity.

Table 7: Estimation results

Variable	coefficient	Standard Error	t value	p-value
P_l	-0.24	0.07	-3.38	0.000
$D_{front,l}$	-1.94	0.08	-22.12	0.000
$D_{center,l}$	-0.96	0.06	-16.92	0.000
$Aisle_l$	5.76	1.27	4.55	0.000
$S_{2,l}$	-4.03	1.00	-4.04	0.000
$S_{3,l}$	25.66	0.96	26.61	0.000
$S_{4,l}$	26.75	1.00	26.73	0.000
$S_{5,l}$	28.84	0.97	29.66	0.000
Constant Term	93.82	3.35	27.97	0.000
			R-Squared	0.615

To summarize impacts of seat locations, we found that although whether seats are next to aisles plays some roles on demand, the effects of the distance from the front (or from the stage) and the distance from the center are more important. Thus, in the next section, we develop a two-dimensional "Scaling the House" model where premium seats are located in the front-center area of the venue. Also, we have seen from the empirical results that customers are more sensitive to the distance from the front than the distance from the center, and it is more common for venue managers to divide seats into zones from front to back. We also develop an alternative one-dimensional model which focuses on the impact of the distance from the front for the scaling decisions.

3.4 Models

From empirical results, we have explored how the demand of seats is affected by seat location and prices. In this section, we integrate the effects of price, distance from the front and distance from the center on the ticket demand, and develop models for the optimal one- and two-dimensional "Scaling the house" decisions.

3.4.1 Two-Dimensional "Scaling the House" Model

In this section, we present the two-dimensional zoning model where zones are divided in two directions: (1) from the front to the back, or "row cut" and (2) from the center to the left/right, or "column cut". We consider the following demand function:

$$\lambda(P_i, P_{j,j\neq i}, F, C) = \alpha - \beta_i P_i + \sum_{j\neq i} \beta_{ji} P_j - \beta_F F - \beta_C C + \epsilon, \tag{6}$$

where $i, j \in \{1, 2\}, j \neq i$, and ϵ is the error term. The demand function $\lambda(P_i, P_{j,j\neq i}, F, C)$ is defined as the ticket demand of seat located at distance F from the front and at distance C from the center, when it is priced at $P_i, i \in \{1, 2\}$ (and $P_j, j \neq i$ is the other seating section's price). This demand function is not only sensitive to its own price but can also be sensitive to the prices of other sections, if there is more than one section available. Parameters β_i and β_{ji} denote the price effect and cross section effect of the price of zone j on the demand of zone i, respectively. We include substitution effects since customers may consider switching from one seating section to another when making a purchasing decision. In addition, as we have found in Section 3.3, demand of seats is sensitive to the distance from the stage (F), where β_F represents distance from the front sensitivity, and it is also sensitive to the distance from the center (C), where β_C represents distance from the center sensitivity.

We have observed that the ticket demand decreases with the distance from the front and the distance from the center. In other words, if tickets are offered at the same price, audiences prefer seats in the front/center, rather than the ones located in the back/left or right of the venue. In this case, it is rational to have the premium seating section in those popular locations. Thus, we consider the two-dimensional optimal "Scaling the house" decision where the zoning sections is defined by the row cut (from front to back) and column cut (from the center to the left/right). Specifically, the decision variables are: 1) ρ_1, the number of seating rows from the front to be priced at high price P_1 before switching to lower price P_2 in the next row, and 2) η_1, the number of seats from seating rows' center to be priced at high price P_1, before switching for lower price P_2 in the left and right areas.

Define r as the total number of seating rows from the front to the back, and c as the average number of seats from the center to the left (or to the right) of the stage, that is one-half of the number of seats in each row. We use symmetric seating sections since we have seen demand symmetrically decreases from the middle to the left/right of the venue (Figure 17). However, our model can be applied to asymmetric sections (if desired) by starting at the column whose demand is the most and calculate the optimal column to switch from that point. The left diagram in Figure 18 illustrates the two-dimensional seating sections in our model. The right diagram in Figure 18 represents the right half of the venue, where the size of each section is determined via the decision variables $\rho_1 \in [0, r]$ and $\eta_1 \in [0, c]$, respectively.

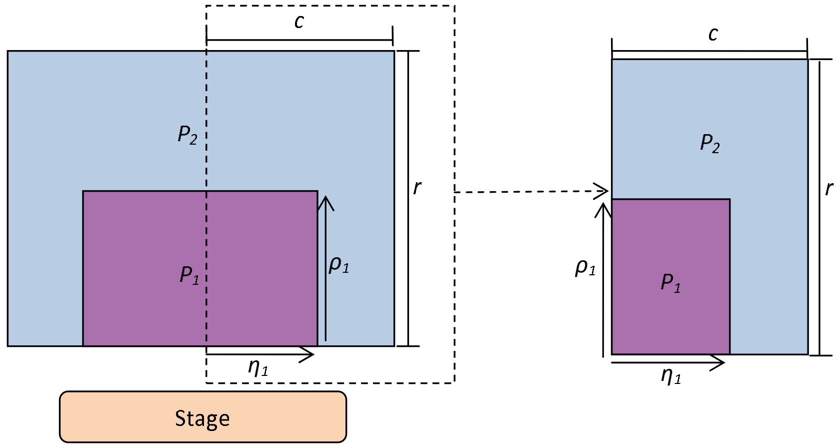

Figure 18: An example of two-dimensional seating sections priced at P_1 and P_2, where $P_1 > P_2$ (left diagram) and the right-half of the venue with decision variables, ρ_1 and η_1 (right diagram)

The key model assumptions are as follows;

1. The entertainment venue manager has predetermined prices, P_1 and P_2, where $0 < P_2 < P_1$.

2. The capacity is fixed and the number of seats is equal to $2 \times c \times r$ (no inventory replenishment).

3. Demand for each seating row is deterministic (as this is a planning decision based on season average) and given by equation (6) with the following characteristic:

i) $\alpha > 0, \beta_i > 0, \beta_{ij} > 0, \beta_F > 0, \beta_C > 0$, where $i, j \in \{1, 2\}, j \neq i$

ii) $\beta_i > \beta_{ij}$, where $i, j \in \{1, 2\}, j \neq i$

This means a unit increase (decrease) in the price causes a greater decrease (increase) in its own seating section's demand than an increase (decrease) in another seating section's demand. This assumption also ensures that the overall market size of the ticket demand decreases in the prices of all seating sections.

iii) $\beta_i > \beta_{ji}$, where $i, j \in \{1, 2\}, j \neq i$

This assumption ensures that the ticket demand generated by a seating zone is affected more by a unit change in its own price than by a unit change in other sections' prices.

4. We assume non-negative demand function and the total demand of seats when priced at either P_1 or P_2 is less than the total capacity. Otherwise, the problem becomes trivial since all seats can be sold at price P_1 or P_2, and the action of "scaling the house" is not necessary.

5. If $\alpha + \beta_1 P_1 + \beta_{21} P_2 > 1$, we assume all seats where $F \leq \frac{\alpha - \beta_1 P_1 + \beta_{21} P_2 - 1}{\beta_F}$ and $C \leq \frac{\alpha - \beta_1 P_1 + \beta_{21} P_2 - 1}{\beta_C}$ can be sold at the high price P_1.

This means if there is excess demand for price P_1 in the popular area, i.e., the front-center location, we assume the excess demand can be captured by the nearby available seats. In other words, all seats in the front-center area with $F \leq M_F$ and $C \leq M_C$ can be sold at price P_1, where M_F and M_C satisfy the following conditions: (1) $2\int_0^{M_C}\int_0^{M_F} E[\lambda(P_1, P_2, F, C)]dFdC = 2M_F M_C$ and (2) $\frac{M_F}{M_C} = \frac{\beta_C}{\beta_F}$. The parameters M_F and M_C in condition (1) represent the seating rows and columns such that the total demand in that area is equal to the seating capacity. Condition (2) states that we assume the excess demand can be captured within the areas depending on how demand is sensitive to the distance from the front and the distance from the center. For instance, if demand is more sensitive to the distance from the front (high β_F), the excess demand will be captured by the available seats with smaller M_F, or closer to the front rather than to the center. Solving (1) and (2), we have $M_F = \frac{\alpha - \beta_1 P_1 + \beta_{21} P_2 - 1}{\beta_F}$ and $M_C = \frac{\alpha - \beta_1 P_1 + \beta_{21} P_2 - 1}{\beta_C}$, respectively.

In this model, we use continuous rows and columns for the zoning decisions, which are made in advance before the season begins, and the same seating chart is applied for all shows of the same performance type (e.g., classical, pop, etc.) during the entire season. In addition, for many venues, pricing decisions are affected by other factors such as economy, competition, etc. Thus, we primarily focus on zoning decisions when the prices are exogenous parameters. We note that this analysis could be repeated with different sets of prices. After determining the optimal "scaling the house" decision for each choice of prices, the venue manager can select the most preferable one.

Define $Rev^{2D}_{n=2}$ as the revenue obtained from the two-dimensional zoning decisions. The venue manager faces the following optimization problem:

$$\begin{aligned}\max_{0\le\rho_1\le r, 0\le\eta_1\le c} Rev^{2D}_{n=2} &= 2\Big(P_1\int_0^{\eta_1}\int_0^{\rho_1} E[\lambda(P_1,P_2,F,C)]dFdC \\ &+ P_2\Big[\int_{\eta_1}^{c}\int_0^{r} E[\lambda(P_2,P_1,F,C)]dFdC \\ &+ \int_0^{\eta_1}\int_{\rho_1}^{r} E[\lambda(P_2,P_1,F,C)]dFdC\Big]\Big), \qquad (7)\end{aligned}$$

subject to:

$$2\int_0^{\eta_1}\int_0^{\rho_1} E[\lambda(P_1,P_2,F,C)]dFdC \le 2\rho_1\eta_1, \quad (8)$$

$$2\int_{\eta_1}^{c}\int_0^{r} E[\lambda(P_2,P_1,F,C)]dFdC + 2\int_0^{\eta_1}\int_{\rho_1}^{r} E[\lambda(P_2,P_1,F,C)]dFdC \le 2cr - 2\rho_1\eta_1. \quad (9)$$

The demand, $\lambda(P_i, P_{j,j\neq i}, F, C)$, where $i,j \in \{1,2\}, j \neq i$, is as defined in equation (6). The objective function (7) is twice the revenue obtained from the right-half of the venue (presented in the right diagram of Figure 18), since we assume symmetry. The revenue of the right diagram in Figure 18 is equal to price P_1 times the number of seats that can be sold at P_1, plus price P_2 times the number of seats that can be sold at P_2. The constraints (8) and (9) ensure that the number of tickets sold does not exceed the capacity of each zone.

Lemma 4. *The two-dimensional zoning model presented in (7) is not jointly concave in the decision variables ρ_1 and η_1.*

We present the proof of this Lemma and other key results of this chapter in the Appendix. Although the revenue function is not jointly concave in the decision variables, we found it is concave in one decision variable given the other is fixed. Consider the second derivatives of the objective function $Rev_{n=2}^{2D}(\rho_1, \eta_1)$ with respect to ρ_1 and η_1, respectively:

$$\frac{\partial^2 Rev_{n=2}^{2D}(\rho_1, \eta_1)}{\partial \rho_1^2} = -2\eta_1 \beta_F (P_1 - P_2) < 0, \tag{10}$$

$$\frac{\partial^2 Rev_{n=2}^{2D}(\rho_1, \eta_1)}{\partial \eta_1^2} = -2\rho_1 \beta_C (P_1 - P_2) < 0. \tag{11}$$

Note that $Rev_{n=2}^{2D}(\rho_1, \eta_1)$ is concave in ρ_1 for a given η_1. Therefore, we can reduce the problem to an optimization problem over a single variable η_1 by first solving for the optimal ρ_1 as a function of η_1, then substituting $\rho_1^*(\eta_1)$ back to $Rev_{n=2}^{2D}(\rho_1^*(\eta_1), \eta_1) = Rev_{n=2}^{2D}(\eta_1)$. Likewise, since $Rev_{n=2}^{2D}(\rho_1, \eta_1)$ is also concave in η_1 for a given ρ_1, we can alternatively start by optimizing η_1 for a given ρ_1 and then substituting $\eta_1^*(\rho_1)$ back to $Rev_{n=2}^{2D}(\rho_1, \eta_1^*(\rho_1)) = Rev_{n=2}^{2D}(\rho_1)$. Both sequential techniques give the same final results and we present the first approach in this chapter.

Lemma 5. *For a fixed η_1, the optimal ρ_1 (the number of seating rows to be price at high price P_1 before switching to price P_2) can be determined uniquely as a function of η_1:*

$$\begin{aligned} \rho_1^*(\eta_1) = {} & \min\{\max\{0, M_F, \\ & \frac{P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) - (P_1 - P_2)\beta_C \frac{\eta_1}{2}}{\beta_F (P_1 - P_2)}\}, r\}, \end{aligned}$$

where $M_F = \frac{\alpha - \beta_1 P_1 + \beta_{21} P_2 - 1}{\beta_F}$.

Lemma 5 provides the optimal ρ_1 as a function of η_1. Then, substituting $\rho_1^*(\eta_1)$ into the revenue function (7), we have the maximization problem over the single variable η_1. We next solve for the optimal η_1^* and substitute back to $\rho_1^*(\eta_1)$, yielding the value of ρ_1^*. The optimal solutions are summarized in the following theorem.

Theorem 4. *Given prices P_1 and P_2, the optimal two-dimensional segmentation with row (ρ_1) and column (η_1) decisions are given below.*

1. *Offer all seats with low price P_2 ($\rho_1^* = 0$ and $\eta_1^* = 0$) when*

$$\begin{aligned} \alpha - \beta_1 P_1 + \beta_{21} P_2 &\leq 1, \textit{and} & (12)\\ P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) &< 0. & (13) \end{aligned}$$

2. *Offer all seats with high price P_1 ($\rho_1^* = r$ and $\eta_1^* = c$) when:*

2.1) if $\beta_C c \leq \beta_F r$,

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F r - \beta_C \frac{c}{2}) > P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F r - \beta_C \frac{c}{2}). \quad (14)$$

2.2) if $\beta_C c > \beta_F r$,

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F \frac{r}{2} - \beta_C c) > P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F \frac{r}{2} - \beta_C c). \quad (15)$$

3. *Offer two seating sections with P_1 and P_2,*

3a) *$\rho_1^* = M_F$ and $\eta_1^* = M_C$ when*

$\alpha - \beta_1 P_1 + \beta_{21} P_2 > 1$, and

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F M_F - \beta_C M_C) < P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F M_F - \beta_C M_C), \quad (16)$$

3b) *$\rho_1^* = \frac{P_1(\alpha - \beta_1 P_1 + \beta_{21}) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) - (P_1 - P_2)(\beta_C \frac{c}{2})}{\beta_F (P_1 - P_2)}$ and $\eta_1^* = c$ when*

$\beta_C c \leq \beta_F r$, and

$$\begin{aligned} (P_1 - P_2)\beta_C \frac{3c}{2} &< P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) & (17)\\ &- P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) \leq (P_1 - P_2)(\beta_F r + \beta_C \frac{c}{2}), \end{aligned}$$

3c) $\rho_1^* = r$ *and* $\eta_1^* = \frac{P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)-(P_1-P_2)\beta_F\frac{r}{2}}{\beta_C(P_1-P_2)}$ *when*

$\beta_C c > \beta_F r$*, and*

$$\begin{aligned}(P_1-P_2)\beta_F\frac{3r}{2} \quad < \quad & P_1(\alpha-\beta_1P_1+\beta_{21}P_2) \\ & - P_2(\alpha-\beta_2P_2+\beta_{12}P_1) \leq (P_1-P_2)(\beta_F\frac{r}{2}+\beta_C c).\end{aligned} \tag{18}$$

3d) $\rho_1^* = \frac{2}{3}[\frac{P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)}{\beta_F(P_1-P_2)}]$ *and*

$\eta_1^* = \frac{2}{3}[\frac{P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)}{\beta_C(P_1-P_2)}]$*, otherwise.*

Theorem 4 provides conditions when it is optimal to offer only one section with price P_2, two sections with price P_1 and P_2, or only one section with price P_1, where the optimal solutions indicate the size of each seating section. While conditions look complicated, there is a reasonable interpretation. From Theorem 4, we can interpret $P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)$ as the higher revenue obtained from the front-center seat ($D = C = 0$) when using price P_1, instead of using price P_2. The conditions that are required for selling all seats with the low price P_2 (case 1) are given by (12) and (13). In other words, if there is no excess demand in the most popular front-center seat for price P_1 (condition (12)), and if it is more beneficial (with higher revenue) to offer price P_2, rather than P_1 for that seat (condition (13)), it will be optimal to use price P_2 for all other less popular seats as well.

On the other hand, in case 2 the venue manager should offer all seats with high price P_1. Note that when $\beta_C c \leq \beta_F r$ (case 2.1), condition (14) implies: if it is more beneficial to sell tickets in the last seating row with high price P_1, then it is most beneficial to use price P_1 for all seats in the venue. When $\beta_C c > \beta_F r$ (case 2.2), condition (15) implies that if more revenue can be obtained on average from pricing at P_1 in the most left/right seating column, then it is optimal to use price P_1 for all seats in the venue.

Next, let us consider the cases outlining two-price sections. From Assumption 5, when there is some excess demand in the most popular front-center seat for price P_1

(i.e., $\alpha - \beta_1 P_1 + \beta_{21} P_2 > 1$), we assume all seats in the front-center area at $D \leq M_F$ and $C \leq M_C$ can be sold at price P_1. Thus, it will not be optimal to switch to the low price P_2 immediately (so, $\rho_1^* > 0$ and $\eta_1^* > 0$). Condition (16) states that suppose it is more beneficial to offer seats at distance M_F and M_C with price P_2, rather than P_1, then it is optimal switch to price P_2 at $\rho_1^* = M_F$ and $\eta_1^* = M_C$ (case 3a).

If $\beta_C c \leq \beta_F r$; and condition (17) holds (i.e, case 3b) then it is optimal to charge the same price for all seats in the same row and to "scale to house" only from front to back (with a row cut). Rearranging the optimal decisions in case 3b, we have $P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F \rho_1^* - \beta_C \frac{c}{2}) = P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F \rho_1^* - \beta_C \frac{c}{2})$. It means the optimal row to switch price is the one whose expected revenue when seats in that row are priced at P_1 is equal to the expected revenue when they are priced at P_2.

If $\beta_C c > \beta_F r$ and condition (18) holds (i.e, case 3c), then it is optimal to "scale to house" only from center to left/right (with column cut). Similarly, the optimal decisions in case 3d imply $P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F \frac{r}{2} - \beta_C \eta_1^*) = P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F \frac{r}{2} - \beta_C \eta_1^*)$; i.e., the optimal seating column to switch price is the one whose expected revenue when seats in that seating column are priced at P_1 is equal to the expected revenue when they are priced at P_2. Otherwise, it is optimal to perform seating segmentations with the solution provided in case 3d.

To provide some visualizations, Figure 19 illustrates examples of the seating sections for each case presented in Theorem 4.

Observation 8. *From Theorem 4:*

i) if $\beta_C c \leq \beta_F r$, the seating chart option E in Figure 19 will never be the optimal solution,

ii) if $\beta_C c > \beta_F r$, the seating chart in option D in Figure 19 will never be the optimal solution.

From Observation 8, combinations of distance sensitivity and venue size can imply

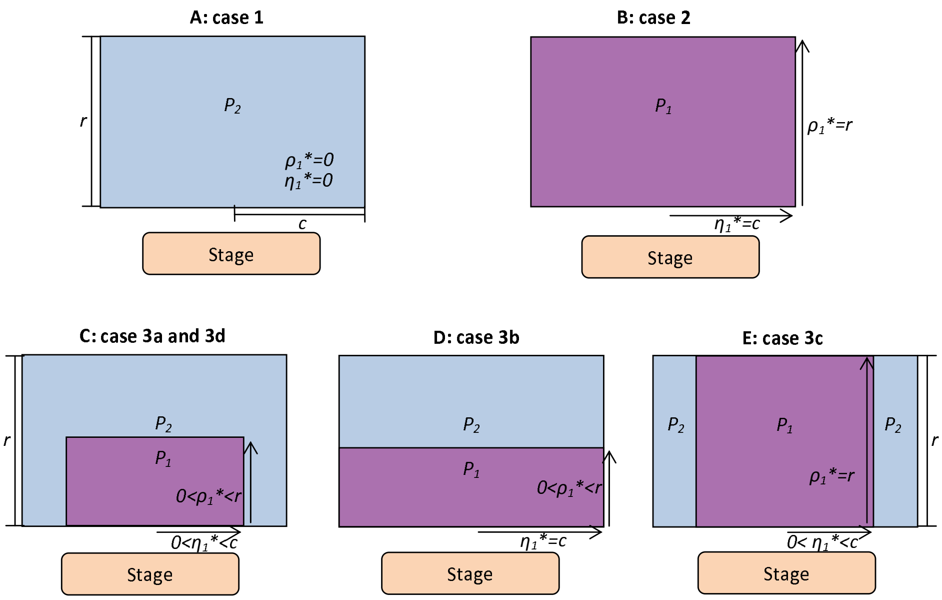

Figure 19: The depictions of the optimal two-dimensional seating sections for each case presented in Theorem 4

insights for venue segmentations. For instance, we found that if demand is very sensitive to the distance from the front (high β_F) and there is a higher number of seating rows than the number of seats from the center ($r > c$), then the venue manager should not choose to have seating segmentations similar to option E in Figure 19. When both β_F and r are high, it is more beneficial to offer the low price in the back rather than on the far sides.

On the other hand, if $\beta_C c > \beta_F r$, then the seating chart option D will never be optimal for the venue since it is better to provide low prices in the left/right seats than in the back seats in this case. In practice, demand is usually more sensitive to the distance from the front than to the distance from the center. This can be why seats in most venues are usually divided into zones from front to back (option D), rather than from center to left/right (option E). Next, we examine the relative size of the optimal solutions for option C of Figure 19 in the following observation.

Observation 9. *If $0 < \rho_1 < r$ and $0 < \eta_1 < c$, then the ratio of the optimal decisions,*

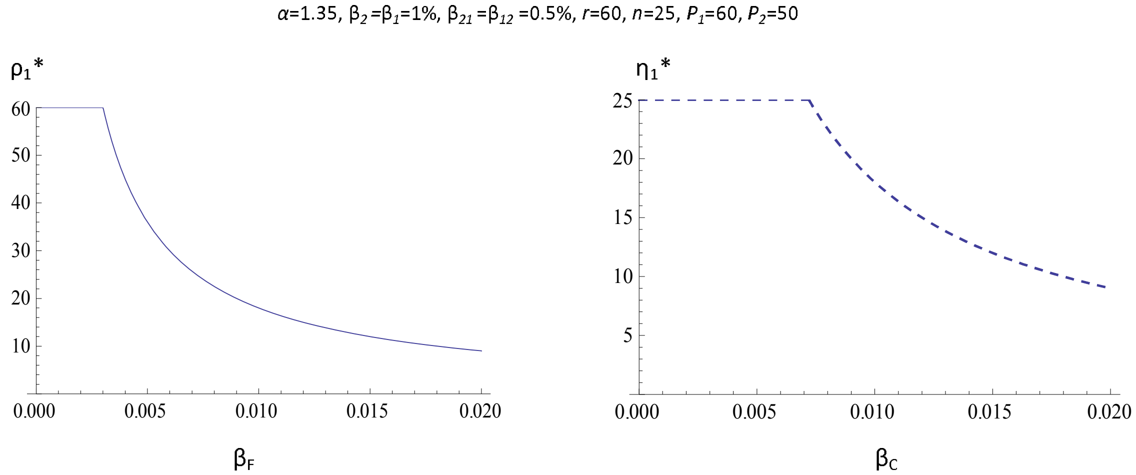

Figure 20: The optimal ρ_1^* at different distance from the front sensitivity, β_F (left), and the optimal η_1^* at different distance from the center sensitivity, β_C (right)

$\frac{\rho_1^*}{\eta_1^*}$, *equals the ratio between the distance from the center sensitivity and the distance from the front sensitivity,* $\frac{\beta_C}{\beta_F}$.

The proof of Observation 9 follows directly from the optimal decisions given in Theorem 4 case 2a and 2b, i.e., $\frac{\rho_1^*}{\eta_1^*} = \frac{\beta_C}{\beta_F}$. It suggests that *the relationship between the distance from the center sensitivity and the distance from the front sensitivity can imply the relative size of the optimal premium section area.* For instance, the premium zone, with high price P_1, should be wide (i.e., small $\frac{\rho_1^*}{\eta_1^*}$)) when customers are relatively less sensitive to the distance from the center as compared to the distance from the front (small $\frac{\beta_C}{\beta_F}$), and vice versa.

Next, we observe how the distance sensitivity impacts each optimal decision individually. Figure 20 (left) depicts the optimal row cut ρ_1^* at different β_F values. As shown in the graph, for small β_F, the optimal ρ_1^* reaches its limit at the last row of the venue (r=60). This is rational since when demand does not significantly decreases as the seats are further from the stage, it is unnecessary to offer the back seats with a low price. *As β_F increases (customers may not want to pay a high price for back seats), the optimal ρ_1^* decreases, suggesting a bigger area for the low-price section in* *the back.* Similar observations hold about the impact of the distance from the center (β_C) on the optimal η_1^*, as shown in Figure 20 (right).

Figure 21 summarizes the impact of both distance sensitivities simultaneously on the optimal zoning decisions (with the same set of numerical parameters as in Figure 20). In this example, $\alpha - \beta_1 P_1 + \beta_{21} P_2 \leq 1$ and $P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) < P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2)$, so we know that case 1 and 2a in Theorem 4 are not optimal. From Figure 21, when both β_F and β_C are low, the ticket demand is not sensitive to the locations of seats and it is optimal to charge all seats with the high price P_1 (case 3). When β_F is low but β_C is high, i.e., the distance from the center is more important than the distance from the front, the optimal solution suggests reducing price only in the left and right areas (case 2b). In contrast, for high β_F and low β_C, i.e., the impact of distance from the front is more important, we found it is optimal to reduce price in the back seats instead of the left/right seats (case 2c). When both β_F and β_C are high, the venue managers should consider offering low price tickets in both areas in the back and the left/right (case 2d). Our results suggest that *different combinations of the distance from the front and distance from the center sensitivities can imply different optimal seating segmentations.*

If there are many types of performances (such as orchestra concert, pop concert and comedian talk show) taking place in the same venue, each type may lead to different distance sensitivities on demand. For instance, pop concerts' demand may be more sensitive to both distances from the front and from the center, since most attendants may prefer to be close to the famous singer (as compared to the demand of broadway shows which may utilize the full stage so the distance from the front sensitivity may not be high). Orchestra's audiences may also have different distance sensitivity compared to comedian talk shows' audiences since the number of performers and shows structures are not the same. Therefore, *it can be beneficial to capture the distance sensitivities of audiences across a category of shows and to offer different*

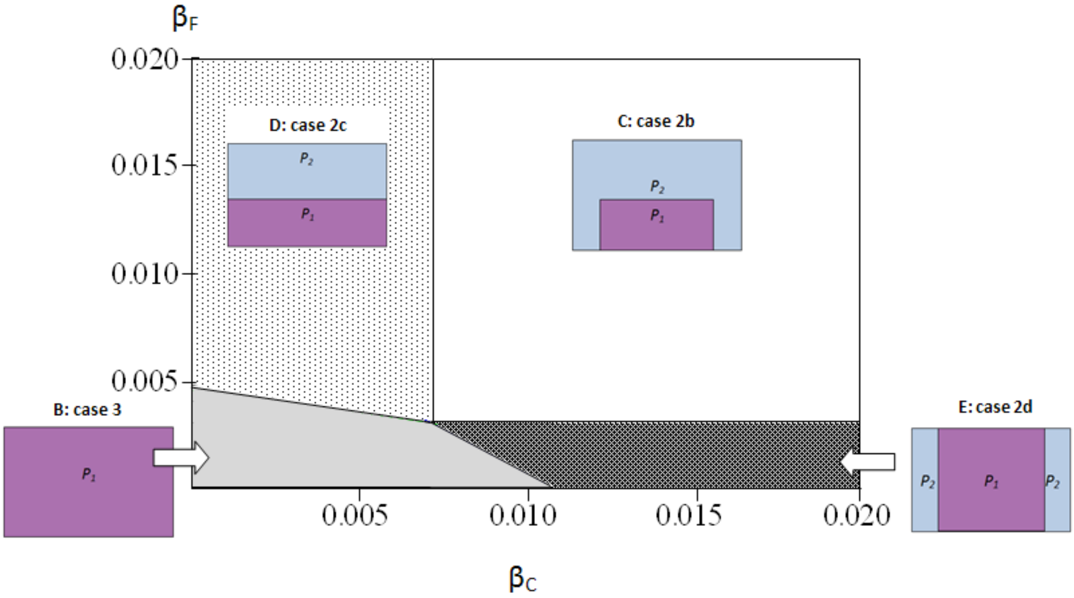

Figure 21: The optimal seating zones described in Theorem 4 at different distance from the front sensitivity, β_F, and distance from the center sensitivity, β_C

zoning sections for each category of show performed in the same venue.

We have found that when ticket demand is not very sensitive to the distance from the center (β_C is small), it may be unnecessary to offer low price P_2 on the side area of the venue. Similarly, for a narrow venue (i.e., small c), customers may not have different perception for seats in the same row. We consider an alternative one-dimensional zoning model in the next section.

3.4.2 One-Dimensional "Scaling the House" Model

From empirical results in Section 3.3, we found demand is more sensitive to the distance from the front than to the distance from the center. Also, it is more common to divide seats into zones from front to back, rather than from center to left/right. In this section, we simplify the previous model to consider one-dimensional zoning decisions defined by row cuts. We consider the following demand function:

$$Q(P_i, P_{j,j\neq i}, F) = a - b_i P_i + \sum_{j\neq i} b_{ji} P_j - b_F F + \epsilon, \quad (19)$$

where $i, j \in \{1, \ldots, n\}, j \neq i$, n is the number of seating sections (where n can be greater than 2), and ϵ is the error term. The demand function $Q(P_i, P_{j,j\neq i}, F)$ is defined as the ticket demand of the seating row located at distance F from the stage when it is priced at $P_i, i \in \{1, \ldots, n\}$ and seats in other sections are sold at price $P_j, j \neq i$. Parameters b_i and b_{ji} denote the price effect and the cross section effect of zone j's price on zone i's demand, respectively. Demand of seats is also sensitive to distance from the stage (F), where b_F represents distance sensitivity. Note that the demand function in (19) is different than the previous section since it represents demand of each seating row; while the demand function given in (6) represents demand of each seat.

Given the demand function described, we next consider how the entertainment venue managers should categorize seats into zones to maximize revenue. In the one-dimensional zoning model, the decision variable is defined by the row cut. The decision variables r_i ($i \in \{1, \ldots, n-1\}$) denote the last seating row to be sold at price P_i before switching to a lower price P_{i+1}. We note that the decision could be column instead of row, if desired, where the optimal solution will be the number of column from the center of rows for a high price before switching to a lower price.

Denote row_c as the average row capacity and r as the total number of rows for the zoning decision. The key model assumptions are as follow:

A. The predetermined price for each section i is P_i ($i \in \{1, \ldots, n\}$) and prices are ordered with subscript i ($P_{i+1} < P_i$ for $i \in \{1, \ldots, n-1\}$).

B. The number of seats in each row is fixed.

C. Demand for each seating row is a known function of parameters and given by (19):

i) $a > 0, b_i > 0, b_{ij} > 0, b_F > 0, P_i > 0$, where $i, j \in \{1, \ldots, n\}, j \neq i$,

ii) $b_i > \sum_{j \neq i} b_{ij}$, where $i, j \in \{1, \ldots, n\}, j \neq i$,

iii) $b_i > \sum_{j \neq i} b_{ji}$, where $i, j \in \{1, \ldots, n\}, j \neq i$,

iv) $0 < \int_0^r E[Q(P_i, P_{j,j \neq i}, F)]dF \leq r \times row_c$, where $i, j \in \{1, \ldots, n\}, j \neq i$.

Note that the justifications of assumptions above are similar to those in Section 3.4.1.

Here, the venue managers face the following expected revenue maximization problem.

$$\max_{0 \leq r_1, \ldots, r_{n-1} \leq r} Rev^{1D} = \sum_{i=1}^{n} P_i \int_{r_{i-1}}^{r_i} E[Q(P_i, P_{j,j \neq i}, F)]dF, \tag{20}$$

subject to:

$$\int_{r_{i-1}}^{r_i} E[Q(P_i, P_{j,j \neq i}, F)]dF \quad \leq \quad (r_i - r_{i-1}) \times row_c, \qquad i \in \{1, \ldots, n\}, \tag{21}$$

where $r_0 = 0, r_n = r$. The demand, $Q(P_i, P_{j,j \neq i}, F)$, is as defined in equation (19). The objective function (20) is the total revenue obtained from each show, which is the summation of the price of each section times the expected total number of seats that can be sold in that section. The capacity constraint (21) ensures the number of tickets sold does not exceed the capacity of each zone. We present the structure of the described problem in the following Lemma.

Lemma 6. *The revenue maximization problem given in (20) and (21), is jointly concave in the zoning decision variables* $r_i, \forall i \in \{1, \ldots, n-1\}$.

The concavity result given by Lemma 6 guarantees a global maximum expected revenue, achieved by the optimal zoning decisions across multiple sections. We can find the solutions by setting the first order condition of each variable to zero and solving simultaneously. Thus, the one-dimensional model has a computational advantage over the two-dimensional problem presented in section 3.4.1.

To demonstrate further structure of model solutions, let us consider the case when $n = 2$, where P_1 and P_2 denote the ticket prices of the first and the second zones, respectively ($P_1 > P_2$). The decision variable, r_1, represents the last row of seats to be sold at price P_1, before switching to price P_2. The revenue maximization problem in (20) and (21) reduces to

$$\max_{0 \leq r_1 \leq r} Rev^{1D}_{n=2} = P_1 \int_0^{r_1} (a - b_1 P_1 + b_{21} P_2 - b_F F) dF + P_2 \int_{r_1}^{r} (a - b_2 P_2 + b_{12} P_1 - b_F F) dF, \tag{22}$$

subject to:

$$\int_0^{r_1} (a - b_1 P_1 + b_{21} P_2 - b_F F) dF \leq r_1 \times row_c, \tag{23}$$

$$\int_{r_1}^{r} (a - b_2 P_2 + b_{12} P_1 - b_F F) dF \leq (r - r_1) \times row_c. \tag{24}$$

The optimal solution is identified in the following Theorem.

Theorem 5. *The optimal row (r_1^*) to be priced at the high price P_1 before switching to the lower price P_2, is given by:*

Case 1. *if* $a - b_1 P_1 + b_{21} P_2 \leq row_c$,

i) $r_1^* = 0$ *when* $P_2(a - b_2 P_2 + b_{12} P_1) > P_1(a - b_1 P_1 + b_{21} P_2)$,

ii) $r_1^* = r$ *when* $P_1(a - b_1 P_1 + b_{21} P_2 - b_F r) > P_2(a - b_2 P_2 + b_{12} P_1 - b_F r)$,

iii) $r_1^* = \frac{a(P_1 - P_2) - b_1 P_1^2 + b_2 P_2^2 - P_1 P_2 (b_{12} - b_{21})}{b_F (P_1 - P_2)}$, *otherwise.*

Case 2. *if* $a - b_1 P_1 + b_{21} P_2 > row_c$,

i) $r_1^* = m = \frac{2(a - b_1 P_1 + b_{21} P_2 - row_c)}{b_F}$, *when* $P_2(a - b_2 P_2 + b_{12} P_1 - b_F m) > P_1(a - b_1 P_1 + b_{21} P_2 - b_F m)$,

ii) $r_1^* = r$ *when* $P_1(a - b_1 P_1 + b_{21} P_2 - b_F r) > P_2(a - b_2 P_2 + b_{12} P_1 - b_F r)$,

iii) $r_1^* = \frac{a(P_1 - P_2) - b_1 P_1^2 + b_2 P_2^2 - P_1 P_2 (b_{12} - b_{21})}{b_F (P_1 - P_2)}$, *otherwise.*

Let us examine the parameter condition of each solution case for insights. The only difference between case 1 and case 2 of Theorem 5 is that we expect demand in the first row when charged at price P_1 to exceed the row capacity row_c, i.e., $a - b_1P_1 + b_{21}P_2 > row_c$ in case 2. Therefore, there is no incentive to switch to the lower price P_2 in the first row since all seats in that row can be sold at the higher price P_1, leading to a higher revenue. In case 1(i), $a - b_1P_1 + b_{21}P_2 \leq row_c$ and $P_2(a - b_2P_2 + b_{12}P_1) > P_1(a - b_1P_1 + b_{21}P_2)$, implying the first seating row's revenue when priced at the lower price P_2 is greater than the revenue when priced at the higher price P_1. In this case, the revenue function is decreasing in $r_1 \in [0, r]$, so it is optimal to switch to price P_2 immediately at $r_1^* = 0$. This can occur when the expected demand is very low and/or when the demand is very sensitive to the high price P_1. Thus, the higher revenue gained from charging a higher price cannot offset the loss of revenue from the decreased demand.

On the other hand, in case 1(ii) and 2(ii), all seats should be priced at the high price P_1 if the last seating row's revenue when priced at the higher price P_1 is greater than the revenue when priced at the low price P_2 (i.e., $P_1(a - b_1P_1 + b_{21}P_2 - b_Fr) > P_2(a - b_2P_2 + b_{12}P_1 - b_Fr)$). It means the revenue function is increasing in $r_1 \in [0, r]$ and the optimal $r_1^* = r$. This can happen when the ticket demand is not very sensitive to price, so it may be unnecessary to attract more demand with low price tickets.

Cases 1(iii) and 2(iii) are what we expect to happen more frequently in practice, which is that the venue managers should offer both prices, P_1 and P_2. *The optimal zoning decision occurs at row r_1^*, in which the expected revenue when seats in that row are priced at P_1 is equal to the expected revenue when they are priced at P_2.* This row can be referred to as the "*indifference*" row, whose revenues are the same for either price. The interpretations of each solution case from Theorem 5 provide easy managerial insights on how the venue management should make the optimal "scaling the house" decisions.

In the following Lemmas, we establish some comparative statics for the optimal zoning decision for two sections, given in Theorem 5 to explore the behavior of the optimal r_1^* as marginal model parameters change.

Lemma 7. *As the marginal price sensitivity, $b_1(b_2)$, increases, the optimal row to switch from zone 1 to zone 2, i.e., price P_1 to P_2, should be moved to be closer to (further from) the front.*

Lemma 7 implies that the more customers are sensitive to price $P_1(P_2)$, the fewer the number of rows that should be priced at $P_1(P_2)$. For instance, if the venue management has realized that demand will decrease more dramatically with high prices (b_1 is larger), then fewer rows should be priced at P_1 so that they can encourage higher revenue from a greater proportion of seats charged at the lower price, P_2.

Lemma 8. *As the marginal price substitution effect, $b_{12}(b_{21})$, increases, the optimal row to switch from price P_1 to price P_2 should be closer to (further from) the front.*

The cross section sensitivity captures how demand changes when prices of other seating sections changes. When b_{12} is larger, it means customers of the back section become more sensitive to the price of the front section; e.g., they are likely to switch from the low-price zone if the premium-zone price P_1 is decreased. In this case, we found the venue managers should shift the low-price section closer to the stage (decrease r_1^*) for higher overall revenue.

On the other hand, when b_{21} becomes larger, customers of the front section (with high price P_1) become more sensitive to price P_2 of the back section; e.g., customers are likely to switch from the front to the back section for a lower price. Then, from Lemma 8, we should shift the low-price section further from the stage and offer more seating rows for the high price section. A reason is when customers are considering to switch to pay a lower price, they will find those low-price seats are further from the stage. Note that this Lemmas hold with more sections as well.

Lemma 9. *The optimal decision, r_1^*, decreases with the distance sensitivity, b_F.*

Lemma 9 states that As the marginal distance sensitivity of demand (b_F) increases, i.e., the demand decreases more significantly with the distance from the stage, the optimal row to switch from price P_1 to price P_2 should be moved closer to the front. An explanation is that when demand significantly decreases with distance, people may not want to pay a high price for seats located further in the back of the venue. Thus, when b_F is high, the venue managers should consider having bigger low-price zones to induce more sales and maximize the overall revenue.

Lemma 10. *The optimal zoning decision, r_1^*, is non-monotonic with ticket prices P_1 and P_2. Specifically, it depends on the following conditions:*

i) As the marginal price P_1 increases, the optimal row to switch from price P_1 to price P_2 should be moved closer to the front if $b_1(P_1)^2 + b_2(P_2)^2 + b_{21}(P_2)^2 > 2b_1P_1P_2 + b_{12}(P_2)^2$. Otherwise, it should be moved closer to the back.

ii) As the marginal price P_2 increases, the optimal row to switch from price P_1 to price P_2 should be moved closer to the front if $b_1(P_1)^2 + b_2(P_2)^2 + b_{12}(P_1)^2 > 2b_2P_1P_2 + b_{21}(P_1)^2$. Otherwise, it should be moved closer to the back.

From Lemma 10, unlike other model parameters such as b_1, b_2, b_{21}, b_{12} and b_F, we found P_1 and P_2 can have different effects on the optimal solution. For example, if the exogenous price parameters are higher, the optimal zoning decisions may suggest to either increase or decrease size of the premium zone (depending on the conditions given above). Table 8 summarizes the comparative statics of the optimal decision r_1^* and the optimal revenue $Rev_{n=2}^*$ to the model parameters, when $r_1^* \in (max\{0, m\}, r)$. The cases where the behavior of the optimal zoning decision (and the revenue) changes conditionally are indicated by a symbol "E", stating it can go either way, depending on the conditions of parameters.

Table 8: Summary of comparative statics on the optimal zoning decision and the optimal revenue

	r_1^*	$Rev_{n=2}^{*1D}$
$a \nearrow$	$\nearrow$	$\nearrow$
$b_1 \nearrow$	$\searrow$	$\searrow$
$b_2 \nearrow$	$\nearrow$	$\searrow$
$b_{12} \nearrow$	$\searrow$	$\nearrow$
$b_{21} \nearrow$	$\nearrow$	$\nearrow$
$b_F \nearrow$	$\searrow$	$\searrow$
$P_1 \nearrow$	E	E
$P_2 \nearrow$	E	E

In real life, it is possible that some model parameters can be misestimated, leading to a lower revenue than the true optimal. We discuss this issue in the next section.

3.5 Robustness of the Models

In this section, we consider the robustness of both two-dimensional and one-dimensional models to estimation errors in the price sensitivity and the two distance sensitivity values. The goal of this section is to examine an aspect of model performance by quantifying the percentage profit loss due to misestimating model parameters. This can generate insights on which parameter estimation's accuracy is more critical in achieving the highest possible revenue, for example.

In the analysis, we start by computing the optimal seating sections at each misestimated value of the parameter and the optimal seating sections with the true value. Next, we calculate revenues from the untrue solutions, compared to the actual optimal revenue (where the accurate estimate was used). For simplicity, we assume the price sensitivities of each section are equal ($\beta_1 = \beta_2$ for the first and $b_1 = b_2$ for the second model). Figure 22 shows the percent revenue loss of the two-dimensional model caused by the estimation errors in (a) the price sensitivity, β_1, (b) the distance from the front sensitivity, β_F and (c) the distance from the center sensitivity, β_C, respectively. For the one-dimensional model, Figure 23 presents a similar analysis with respect to errors in (a) the price sensitivity, b_1 and (b) the distance sensitivity, b_F.

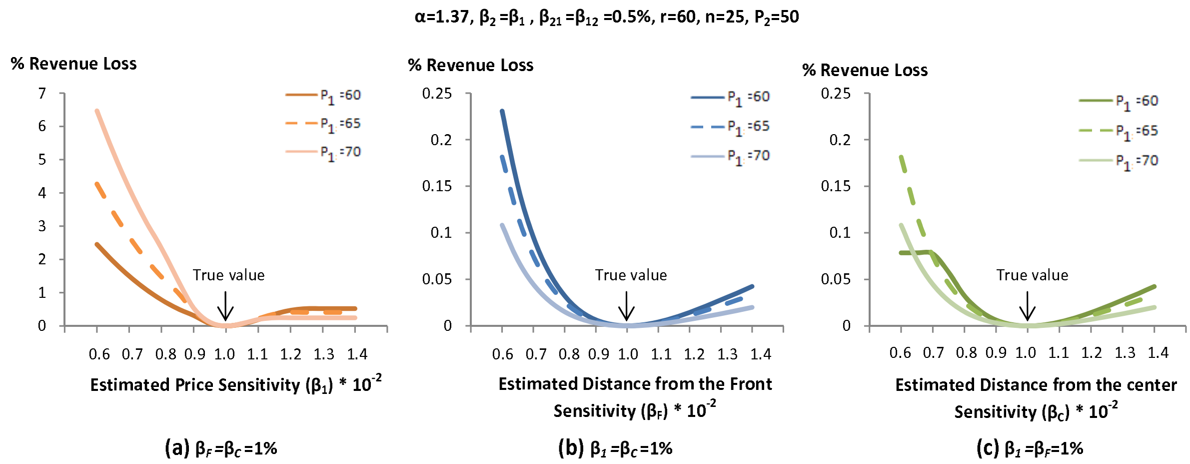

Figure 22: Percent revenue loss of the two-dimensional model caused by estimation errors in (a) β_1, (b) β_F and (c) β_C

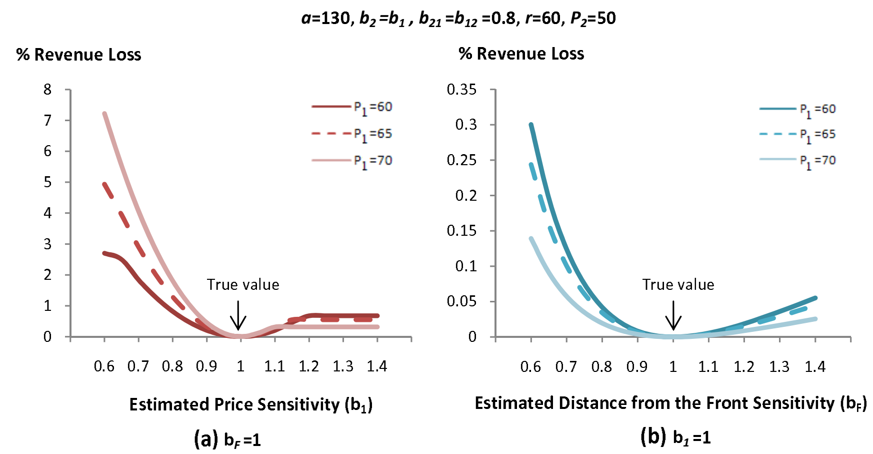

Figure 23: Percent revenue loss of the one-dimensional model caused by the estimation errors in (a) b_1 and (b) b_F

Observation 10. *For both models:*

i) Underestimation of the model parameters causes greater revenue losses than overestimation;

ii) Estimation errors in the price sensitivity lead to higher revenue losses than errors in the distance sensitivity values.

When the price sensitivity is underestimated, both models choose to offer more seats at the high price, leading to a significant drop in ticket demand and overall revenue. On the other hand, when the venue managers underestimate the price sensitivity, the optimal zone decisions will offer more seats with the low price. The expected demand is higher than the optimal but the revenue is less. Likewise, the same reasoning applies for the misestimation of the distance from the front sensitivity and the distance from the center sensitivity. We found that overestimation errors lead to higher revenue losses than underestimation for both price and distance sensitivities. The results imply *a risk-averse venue manager who tends to overestimate sensitivity of demand may achieve higher revenue than a risk-taker manager who underestimates them.*

When we compare revenue losses, the estimation errors from the price sensitivity are more costly than those from the distance sensitivities for both models. A possible reason is that the gap between the zoning decisions and the actual optimal decisions is greater when the price sensitivity is misestimated than when the distance sensitivities are misestimated. *This finding emphasizes the importance of correctly estimating price sensitivity parameters to achieve the maximum revenue in this ticket industry.*

3.6 Comparisons among Different "Scaling the House" Models

We have presented the two-dimensional and the one-dimensional scaling the house models in Sections 3.4.1 and 3.4.2, respectively. In this section, we conduct various computational experiments to answer the following questions; (1) for the one-dimensional scaling the house, when is a venue suitable for sectioning from the front to the back (row cut), rather than sectioning from the center to the left/right (column cut), and vice versa?; (2) when is it beneficial to perform two-dimensional scaling the house (both row and column cuts) instead of a simple one-dimensional sectioning? These insights can help venue managers make more effective decisions when it comes to model selection.

Next, we explore what happens if the venue management divides seats into zones from the front to the back (row cut), rather than from the center to the left/right (column cut) when $\beta_F r > \beta_C c$. Figure 24 shows the percent benefit gained from implementing the row cut instead of the column cut at different values of distance from the front sensitivity (β_F) and the number of rows (r). We can see that as β_F increases, the percent revenue gained from using row cuts (in place of column cuts) also increases. In addition, sectioning seats by row cut is more advantageous when r is high, e.g., the percent increase in revenue is the most when r=55 for all values of β_F. The results imply that *if only the one-dimensional cuts are considered, the larger the number of rows and the higher the distance from the front sensitivity, the more attractive it is to consider zoning from the front to the back, rather than from the center to the left/right.*

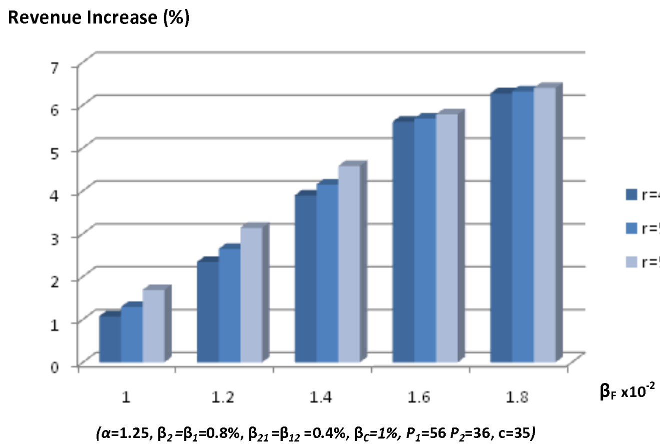

Figure 24: The percentage revenue increase from implementing the one-dimensional row cut instead of the one-dimensional column cut at different values of β_F and r

In contrast, from Theorem 4 in section 3.4.1, it is never be the optimal to perform row cuts when $\beta_F r < \beta_C c$. Figure 25 depicts the percent increase in revenue from using the column cut in place of the row cut, at different values of the center sensitivity (β_C) and number of seats from the center to the left/right (c). Similarly, we can see from Figure 25 that when β_C and/or c increases, the percent revenue gained from using the column cut (instead of the row cut) also increases. Therefore, by looking at the distance sensitivity parameters and the structure of the venue (i.e., number of rows and seats), it is possible to determine which type of one-dimensional sections is more suitable. *The higher values of β_C and c (or, β_F and r), the higher tendency that the one-dimensional "Scaling the House" from the center to the left/right (or, from the front to the back) is more suitable for the venue.*

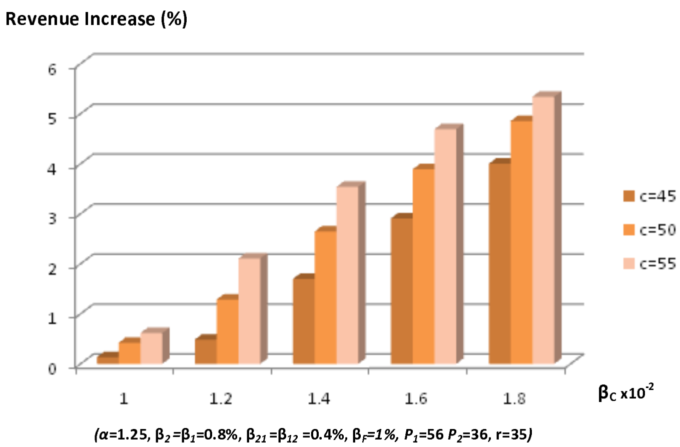

Figure 25: The percentage revenue increase from implementing the one-dimensional column cut instead of the one-dimensional row cut at different values of β_C and c (when $\beta_F r > \beta_C c$)

The next interesting question is; when is it beneficial to use the two-dimensional cuts (both row and column as presented in section 3.4.1) instead of a simple one-dimensional model as presented in section 3.4.2? For instance, how does the venue's size impact the benefit gained from performing both row and column cuts, as compared to a simple row cut? Figure 26 shows the percentage increase in revenue when implementing the two-dimensional instead of the one-dimensional sections at different venue's sizes, r and c. We can notice that *as the venue size increases (in both directions), it is more beneficial to perform both row and column cuts. Moreover, for venues with the same size, as the distance sensitivity values (both β_F and β_C) increase, the benefits from using two-dimensional zones also increases.*

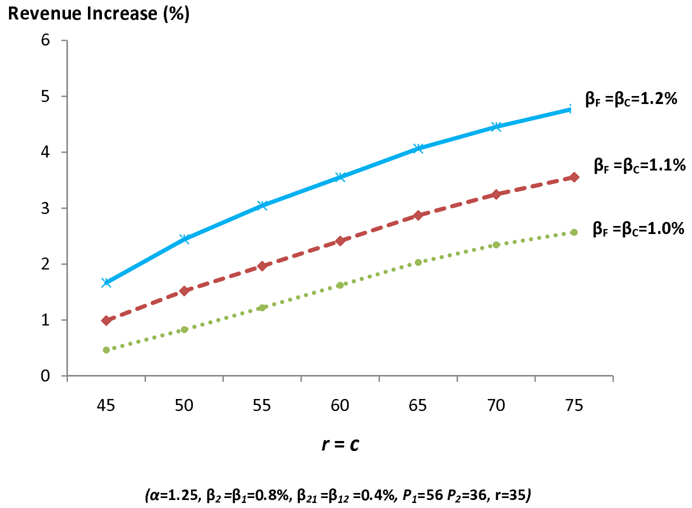

Figure 26: The percentage revenue increase from implementing the two-dimensional row and column cuts instead of the one-dimensional row cut at different values of r and c (when $\beta_F r < \beta_C c$)

In practice, the entertainment venues may differ in term of their shapes. For instance, a venue may have more rows than the number of seats in each row (represented in the left diagram of Figure 27). In contrast, other venues may have more seats in each row than the number of rows (the right diagram of Figure 27). In the next analysis, we explore performances of different zoning models when venues have different shapes (e.g., different $\frac{r}{c}$). Let "2D", "1Drow" and "1Dcol" represent the two dimensional zones, the one-dimensional zones from the front to the back and the one-dimensional zones from the center to the left/right, respectively.

The left diagram of Figure 28 shows the percentage revenue increased when implementing the "2D" instead of: (1) the “1Drow” and (2) the “1Dcol”, where $\beta_F = \beta_C = 1\%$. It can be observed that as $\frac{r}{c}$ increases, i.e., there is a higher number of rows compared to the number of seats in each row (large $\frac{r}{c}$), the revenue gain from using the “2D” zone instead of the “1Drow” zone (“2D vs 1Drow”) decreases. It implies that when $\frac{r}{c}$ is small, the “2D” seating zone performs much better than the “1Drow” seating zone. However, as the number of rows are greater or the number of seats in each row are fewer, it becomes less necessary to consider the “2D” model because its benefits become similar to the “1Drow” model. On the other hand, when $\frac{r}{c}$ is small, the benefit gained from choosing the “2D” zone instead of “1Dcol” zone (“2D vs 1Dcol”) is also small. It implies the “2D” and the “1Dcol” zones are not much different at small $\frac{r}{c}$ and the benefit gained becomes larger as $\frac{r}{c}$ increases.

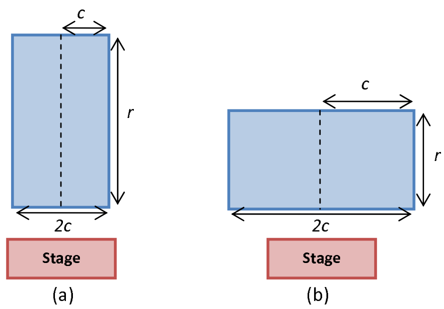

Figure 27: Venue shapes examples

The right diagram of Figure 28 also demonstrates the percentage revenue increased when implementing the "2D" instead of the “1Drow” and the “1Dcol”. While β_F and β_C were equal before, now β_C is twice as big, i.e., $\beta_F = 1\%$ and β_C=2%, or $\beta_F = \frac{1}{2}\beta_C$ in this case. We can see similar trends for both “2D vs 1Drow” and “2D vs 1Dcol” lines in the left and the right graphs. However, a difference is that dividing seats in the “2D” zone is less worthwhile if (1) $\frac{r}{c} << 1$ and $\frac{r}{2c} >> 1$ for the left diagram (where $\beta_F = \beta_C$) and (2) $\frac{r}{c} << 2$ and $\frac{r}{2c} >> 2$ for the right diagram (where $\beta_F = \frac{1}{2}\beta_C$), since there exists a one-dimensional cut yielding quite similar performances. When $\frac{r}{c}$ is close to 1 in Figure 28 (left) and close to 2 in Figure 28 (right), it is more beneficial

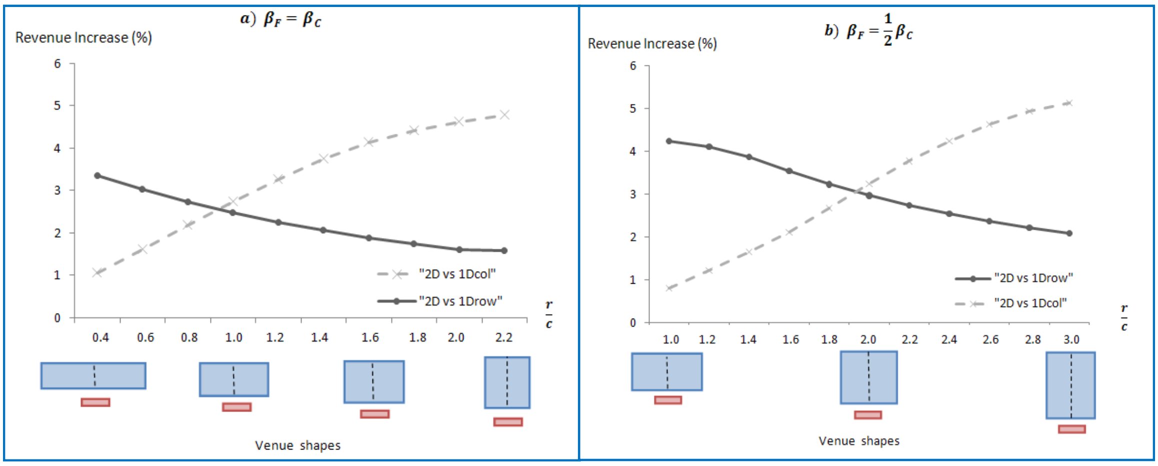

Figure 28: The percentage revenue increase from implementing the two-dimensional row and column cuts instead of: 1) the one-dimensional row cut ("2D vs 1Drow") and 2) the one-dimensional column cut ("2D vs 1Dcol"), for different venue shapes

to apply the " 2D" zone since it gives higher revenue than both "1Drow" and "1Dcol" zoning decisions[1]. In summary, we found *the "2D" model is most worthwhile when* $\frac{r}{c} \approx \frac{\beta_C}{\beta_F}$. Note that if a venue has several seating areas (e.g., different floors), this insight can motivate different zoning models for seating areas with different shapes within same the venue as well.

3.7 Conclusions

Although sectioning seats into zones for different prices ("Scaling the house") is one of the most important decisions in the entertainment industry, systematic approaches for this problem have not been developed so far. The goal of this chapter is, therefore, to develop models for the optimal "Scaling the house" decisions. We obtained transactional level data of over 300 shows from a major performance arts organization in the United States. We empirically explored how seating locations drive ticket demand within the venue and found that *in addition to price, demand is significantly driven*

[1]Parameter values for Figure 28 are $\alpha = 1.25$, $\beta_1 = \beta_2 = 1\%$, $\beta_{12} = \beta_{21} = 0.5\%$, $P_1 = 60$ and $P_2 = 45$.

by the distance from the front and the distance from the center.

Given the characteristics of demand, we developed a two-dimensional "Scaling the House" model and identified the optimal closed form solutions of the number of seating rows (from the front) and the number of seats (from the center of rows) to be charged at a premium price. We found that *for the two dimensional zones, the ratio of the optimal row cut and column cut equals the ratio between distance from the center sensitivity and distance from the front sensitivity.* When ticket demand does not significantly decrease with the distance from the center, the optimal seating zones will be defined only by row cuts, where seats in the same row have the same price. In fact, from our empirical results, customers are more sensitive to the distance from the front than the distance from the center. Thus, we developed an alternative one-dimensional zoning model (with row cuts). We showed *the optimal cutting point occurs at the seating row whose expected revenue when charging the a high price is equal to its expected revenue when charging the a low price.*

To obtain managerial insights, we performed comparative statics of the two-price zoning decisions. We showed that *the optimal number of high-price seating rows decreases with the customers' price sensitivity.* Thus, if customers are very sensitive to high prices, the venue managers should provide a smaller premium section and induce more revenue with more low-priced seats. In addition, *the optimal number of high-price seating rows decreases with the distance from the front sensitivity.* When demand decreases significantly with distance, people may not want to pay a high price for seats further in the back, so the venue managers may consider having a bigger low-price section to motivate more revenue. The result also implies that *when different types of performances (with different distance sensitivity on demand) are performed in the same venue, the management should consider having different seating charts for each show type to maximize revenue.*

In robustness analysis, we found that underestimation of model parameters causes greater revenue losses as compared to overestimation (for both two dimensional and one dimensional zoning models). Therefore, a risk-averse venue manager who tends to overestimate sensitivity of demand may achieve higher revenue than a risk-taker manager who underestimates them. We perform several computational experiments to identify how different venue's characteristics impact the suitable zoning model choices. The results show that it is most worthwhile to use the two-dimensional zoning model, in place of the one-dimensional model when $\frac{r}{c} \approx \frac{\beta_C}{\beta_F}$. Thus, different venue's shape can imply different zoning models and our insights can help the venue managers select appropriate seating sections, based on the structures of their venues.

It is worthwhile to note that our "Scaling the House" models and results also provide some implications to other industries whose demand is sensitive to spacial locations. For example, in the airline industry, customers may prefer some locations of seats in the aircraft such as window/aisle and front/back areas. In fact, airlines (e.g. US Airways, Northwest and Spirit) have started to charge premium prices for some types of popular seats [139]. It would be interesting to identify the optimal zoning solutions for this business. Another example is the online advertising industry, where some locations on the web page are considered to be more valuable for ad placements. For instance, the top-page space may be the most stand-out area for customers' notification, so web site managers can consider "Scaling the Page" for different advertisement fees to maximize their revenue as well. We hope our work in this chapter also motivates the future works of spatial locations impacts on revenue management applications in broader industries.

CHAPTER IV

DYNAMIC PRICING WITH DEMAND LEARNING FOR THE SPORT AND ENTERTAINMENT TICKET INDUSTRY

4.1 Introduction

Revenue management (RM) has attracted much attention and been proven as one of the most effective practices to increase profitability for many industries. RM first emerged in the airline industry in the context of passenger booking problems ([7], [94] and [125]). Then, it has played important roles in improving the performance of many industries, selling fixed perishable capacity with high set-up costs such as hotels ([12], [13] and [91]), cruise lines [82], passenger railways [31], and rental car companies ([25] and [55]). Ultimately, RM is used to support decision making to achieve the goal of selling the right amount of product at the right price to the right customers at the right time [18].

Dynamic pricing is an RM tool that is widely used to manage and control demand at different points of time. Sellers adjust prices to increase or decrease demand in the short run so that it can be matched with their available resources. The main objective of dynamic pricing is to find an optimal dynamic policy to balance utilization of the available capacity so that the revenue can be maximized over the selling period. There are extensive papers exploring a variety of dynamic pricing topics (e.g., [11], [28], [51], [52] [73], [87], [90] and [98]). While dynamic pricing has been intensively studied in the travel industries, the Sport and Entertainment (S&E) industry is another business that has potential to be improved by the idea but still has not received as much attention [41].

There are approximately 1,953 sport stadiums and 236 performance venues in the United States, with the total revenue of 44.2 billion reported in 2009 [119]. Similar to the airline business, the number of the S&E tickets is fixed and the are "perishable" since they have no value after the event date. However, the S&E industry's characteristics differ from airlines or hotels in many ways. First, a much higher percentage of entertainment tickets are purchased on the day of the show than on the day of a flight or on the day of a hotel stay [41]. Secondly, while important factors of demand are date/time of a flight for airlines and day/month of a stay for hotels, the S&E ticket demand is also very related to the sport teams or performance artists' popularity. Moreover, consumer tastes and economic conditions change over time and events vary from one year to another, so it is difficult to incorporate all uncertainties to precisely predict S&E ticket sales. For these reasons, a different pricing model is necessary for the S&E industry.

Although there are a number of studies that are related to ticket pricing (e.g., [35], [38] and [123]) and the topic of price variation in the S&E business (e.g., [89] and [120]), the previous research considered static pricing decisions where ticket prices remain constant for the entire selling horizon. It is generally because in the past, most event tickets were sold with a fixed price (independent of when the tickets were sold) due to the limited inventory tracking devices or ticket changes were done on an ad-hoc basis. However, in recent years, tools such as internet based selling systems have become widely available, providing information for real-time demand observations. From our discussions with a performing arts consulting firm, there has been a significant interest from a number of performance arts organizations for methods to apply dynamic discounts/premiums pricing to more effectively manage demand and increase revenue.

For new products (without past sales information) and/or products whose demand patterns may deviate significantly from past history, demand characterization is difficult [83]. Likewise, the S&E demand can be uncertain, especially for a new show or a sports team with varying performance. So, the seller cannot totally rely on past sales history when predicting demand. Early sales observations can be useful for demand information updates. For example, after the selling time starts, the ticket sellers will have a clearer picture whether the games (or shows) are likely to have high or low sales. In this chapter, we consider stochastic S&E ticket demand and incorporate demand learning with Bayesian updates. A stochastic setting is appropriate to capture real-life situations when the paths of demand over time is difficult to be accurately predicted [18]; while demand learning allows the seller to update his beliefs as uncertainty reveals itself.

In this chapter, we assume ticket demand in each period follows a Poisson distribution, where its rate is affected by three components; (1) the artists/sport teams' popularity, (2) the ticket price and (3) the remaining time until the show date. Including timing effects for the S&E ticket demand is motivated by the actual data. Figure 29 depicts the average number of tickets sold for 108 shows performed in a major performance venue in the U.S. during the 2007-2008 season, where the horizontal axis represents the number of months prior to the show dates. The upper and lower lines represent the average plus and minus one standard deviation, respectively. We can see in Figure 29 that the average number of seats sold each month is dependent on time prior to the show and it generally increases when the time is closer to the event date.

In our demand model, the seller can characterize price effects and timing effects on demand, but he has incomplete information about the event popularity (i.e., the base demand rate). The seller forecasts the base demand rate' s initial estimation, and then uses the observed sales to update his belief about the upcoming period's demand. We develop dynamic pricing models in a discrete-finite-time selling horizon. In the first period, the seller determines the optimal base ticket price. In the following periods,

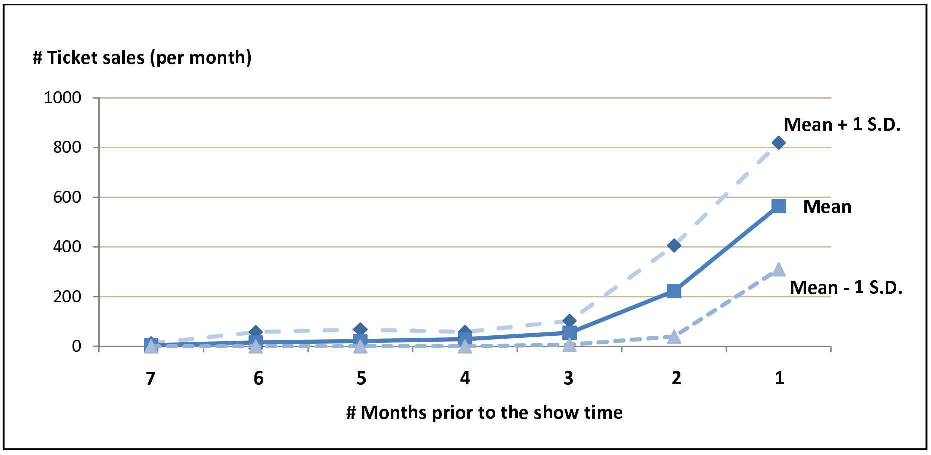

Figure 29: The average ticket sales of 108 shows performed in a major performance venue in the U.S. during the 2007-2008 season, at different times prior to the show

the seller can offer discounts (decrease price) or charge premiums (increase price), with respect to the decided base price. We consider two types of pricing mechanisms. The first model is called "Dynamic discounts/premiums pricing model", where the levels of discounts/premiums can be adjusted in every period. In the case that changing prices too often is difficult or costly, the number of price changes may be restricted. We consider in the second model (called the "Dynamic timing model"), where the seller decides the optimal time to change price and the optimal level of discounts/premiums that will be applied for all remaining periods. These problems are motivated by our discussions on actual practices and methods that are needed in the S&E industry. We address the following research questions:

1. How can the observed sales be used in the demand learning process to improve the forecast, and when is demand learning most beneficial?

2. How are the optimal discounts/premiums related to model parameters such as price sensitivity and remaining inventory of unsold tickets?

3. When is it most worthwhile to adjust discounts/premiums every period and when is it most practical to limit the frequency of price discounts/premiums

changes?

4. If the seller decides to limit the price change to only one, when is it optimal to offer discounts/premiums: early in the selling season, or close to the show date?

This chapter is organized as follows. We begin by presenting a summary of the relevant literature in the next section. In Section 4.3, we discuss demand assumptions and describe how the observed sales are used to update the belief about the upcoming period's demand in the learning process. We next study two dynamic pricing models in the context of stochastic dynamic programs. In Section 4.4, we present the dynamic discounts/premiums pricing model, where the price discounts/premiums adjustment is allowed in every period, and discuss the value of applying demand learning to the pricing model. In Section 4.5, we address an alternative model (i.e., the dynamic timing model) where the number of price changes is limited to only one and present computational experiments comparing the dynamic timing model to the dynamic discounts/premiums pricing model. Finally, Section 4.6 concludes with a summary of insights from the results this chapter.

4.2 Literature Review

This research draws on and contributes to three streams of literature: 1) Sport and Entertainment tickets pricing and timing literature, 2) dynamic pricing for perishable products with no inventory replenishment and 3) demand learning with Bayesian updates. In this section, we review related papers in these three areas.

Most pricing studies in Sport and Entertainment industry are concentrated in economics and marketing fields. An extensive review on pricing practices observed in the entertainment ticket markets can be found in [35]. Rosen and Rosenfield [123] developed an economic model with two seat categories and determined the number of tickets and price for each category that maximizes revenue. Marburger [100] considered a pricing model for performance goods when there are other complementary products sold at the event. All the works above considered deterministic demand with static pricing decisions, i.e., ticket prices remain constant for the entire selling horizon. Jorgensen et al. [70] studied optimal pricing and advertising policies for entertainment events, taking into account diffusion effects and the last-minute market. In their model, the authors assumed if all tickets are not sold by a current period, the remaining tickets will be sold at a reduced price in the last period. Unlike Jorgensen et al. [70], we assume stochastic demand with uncertain arrival rates that can be updated as more demand information is revealed. This topic of pricing with demand learning has not been explored in the S&E ticket pricing literature.

There are a small number of recent papers exploring opportunities for revenue management in the S&E business. Drake et al. [41] considered the optimal time to switch from selling ticket bundles to selling a single ticket. The authors characterized the conditions when only bundles, only individual tickets, or mixed bundling should be offered. While Drake et al. considered a static timing decision that is determined at the beginning of the selling horizon, Yakici et al. [147] extended the previous model to a dynamic switching model. They showed that the optimal time to switch (from selling ticket bundles to selling a single ticket) consists of a set of threshold pairs defined by the remaining inventory and the time left in the horizon. From computational results, they found the profit can be improved by 1 to 2 percent with dynamic switching as compared to static switching. Unlike the papers above, we consider *dynamic pricing models* for the S&E industry in this chapter.

Dynamic pricing has been proven to be an effective tool for revenue management and has been extensively studied in the literature. Bitran and Caldentey [18], and Elmaghraby and Keskinocak [44] provided important reviews on dynamic pricing research. Kincaid and Darling [77] were the first who formally developed a single-product dynamic pricing model as a continuous-time stochastic dynamic program, and they identified important properties of the revenue function. Gallego and van

Ryzin [51] considered a continuous time dynamic pricing problem of a firm selling a fixed stock of items over a finite horizon. The authors derived important structural properties of the optimal solution and developed a deterministic heuristic to solve the problem. In practice, it can be to costly to change prices too often. Feng and Gallego [48] studied pricing models when prices can be changed only once and the second price can be chosen from a set of allowable prices. The authors showed that the optimal time to change price is a function of the remaining time and the remaining inventory level. Bitran and Mondschein [16] considered a dynamic pricing model for seasonal products having stochastic customer arrivals. They found that uncertainty in the demand for new products leads to higher prices, larger discounts, and more unsold inventory. Other dynamic pricing models for limited inventory studies are presented in [17] and [149]. In contrast to these dynamic pricing papers, our work considers demand learning opportunities and explores how the seller should adjust prices as past sales are observed.

In real life, demand learning is possible as selling process evolves and the current inventory level can be observed. Pricing can be an important tool to control not only demand but also the speed of how the seller learns about actual popularity of the products. Bayesian updating is one of the most important and widely used approaches for demand learning. Scarf [126] was the first to introduce a Bayesian approach to inventory models. He formulated a stochastic dynamic program where demand distributions had an unknown parameter but could be updated via Bayes' method. His optimal policy was to order up to a level, where the level was shown to be an increasing function of the cumulative demand in the past periods. Murray and Silver [108] developed an inventory model with a prior purchase probability of a Beta distribution with unknown parameters. After early sales are observed, the distribution was updated to compute the optimal ordering quantity for the next period. Chang and Fyfee [29] studied demand learning in which demand consists of two terms; a fraction of total demand and a noise term, and the distribution of the first part was updated after actual sales were revealed in each period. Eppen and Iyer [45] developed a different learning model for fashion products, where the demand process is defined as a set of pure demand processes having discrete prior distributions, that can be revised as past sales are known. In their model, the seller decides the order quantity and the quantity to be diverted to outlet stores in each period. Other demand learning research in operations management can be found in [3], [21], [50], [61], [65], [84], [96] and [102]. All of these papers assumed price to be exogenous and did not consider the flexibility of a pricing decision. Our model differs from them in that it incorporates demand learning for optimal pricing, rather than for inventory decisions.

There are papers considering both pricing and inventory decisions with the existence of demand learning. Petruzzi and Dada [116] considered optimal pricing and inventory levels under various types of deterministic demand with an unknown function parameter that can be revised after demands in early periods are realized. If the unknown parameter was revealed in any period, then the authors assumed the unknown parameter became certain and did not require learning in any subsequent period. Subrahmanyan and Shoemaker [135] developed a pricing and stocking model when demand had a discrete prior distribution and could be learned using Bayes' rule. Their model's decisions are the optimal price and order quantity for each selling period. Sen and Zhang [127] studied pricing decisions with demand learning via Bayesian updates for style goods industry and considered the value of flexibility of replenishment within the season. In contrast to the papers above, there is no possible inventory replenishment for the S&E tickets during the selling season because it means adding capacity to the venue or arena. So, our work focuses on dynamic pricing decisions without inventory replenishment.

Balvers and Cosimano [4] considered dynamic price adjustment models for a firm selling limited inventory, where demand is a linear function of price with unknown slope and intercept. They described how the continuing process of learning about demand curve affects the pricing decision over time. Aviv and Pazgal [2] considered a partially observable Markov modulated demand model in a discrete time setting for fashion goods. Other dynamic pricing with learning papers that did not consider inventory replenishment include [1], [14] and [92]. Unlike the studies above, timing is important in the S&E industry since ticket demand significantly increases as it is closer to the show date. Therefore, in this chapter we include timing effect on ticket demand. Also we consider an alternative model where ticket price can be changed only once and this change will be applied for all subsequent periods, where we focus on the impacts of different demand pace and demand variation on the timing of price adjustment.

4.3 Demand Learning Model

In this section, we describe our demand model and show how the observed sales are used to update beliefs about demand in the upcoming periods.

We assume the ticket demand in each period t follows a Poisson process with rate $\Theta_t(p_t, \Gamma_t)$, which is affected by three components: (1) the ticket price, (2) the artists/sport teams' popularity and (3) the remaining time until the show date. Note that p_t denotes the ticket price in period t and Γ_t denotes the base demand rate, which represents the expected popularity of the show. The overall demand rate is defined as follows:

$$\Theta_t(p_t, \Gamma_t) = \phi(p_t) g(t) \Gamma_t,$$

where $\phi(p_t)$ represents price effects (i.e., the probability of each arrival purchasing a ticket) and it is decreasing with price p_t. The demand timing effect, $g(t)$, is a decreasing function of t (where $t = n$ at beginning of the selling time and $t = 0$ at the show/event time). This assumption is underlined from the data we observed (e.g., Figure 29) that ticket demand tends to increase closer to the show date.

We note that from statistical theory, the mean and variance of a Poisson distribution are the same [124]. However, we observe that the variance of entertainment tickets is much larger than the mean [1], showing a significant sign of overdispersion, i.e., a greater variance than what is expected in a Poisson distribution. Similar observations are found in other sets of data we have observed. As a result, an assumption that demand follows a general Poisson process with a certain rate is not practical and thus we incorporate uncertainty in our demand function.

From our discussions with the performing arts consulting firm, the show's popularity in customers' perspective is usually uncertain to the seller. Therefore, in our model we assume there is incomplete information on the exact value of the base demand rate, Γ_t. At the beginning of the selling time ($t = n$), we assume Γ_n follows a Gamma distribution with a scale parameter of a and a shape parameter of b. In addition to the show's popularity, a and b may depend on the city in which the show is performed, since we may expect a higher base demand rate for the show taking place in a bigger city. A Gamma distribution allows the flexibility of the shape and position to be changed extensively via parameters a and b, where empirically we found it a good match to past data.

Let M_n be the random demand in period $t = n$. The distribution of M_n, conditional on the base demand rate ($\Gamma_n = \gamma$) and ticket price (p_n), follows a Poisson distribution with rate $\phi(p_n)g(n)\gamma$.

Thus, for demand in the first period, we have:

$$f(M_n = m|\Gamma_n = \gamma, p_n) = \frac{[\phi(p_n)g(n)\gamma]^m \, e^{-\phi(p_n)g(n)\gamma}}{m!}, \forall m \in \{0, 1, 2, \ldots\}. \tag{25}$$

Similarly, the distribution of demand in period t (denoted by M_t), conditional on the

[1] approximately 20-70 times higher than the mean

base demand rate and ticket price, follows a Poisson distribution with rate $\phi(p_t)g(t)\gamma$, i.e.,

$$f(M_t = m|\Gamma_t = \gamma, p_t) = \frac{[\phi(p_t)g(t)\gamma]^m \, e^{-\phi(p_t)g(t)\gamma}}{m!}, \forall m \in \{0,1,2,\ldots\}, \forall t \in \{n-1,\ldots,1\}. \tag{26}$$

Next, we identify the prior distribution of the ticket demand in the first selling period ($t = n$). From Bayes' rule, we have that the distribution of demand (when the ticket price is p_n), unconditional on Γ_n, is given by:

$$\begin{aligned} f(M_n = m|p_n) &= \int_0^\infty f(M_n = m|\Gamma_n = \gamma, p_n) f(\gamma) d\gamma \\ &= \int_0^\infty \left(\frac{[\phi(p_n)g(n)\gamma]^m \, e^{-\phi(p_n)g(n)\gamma}}{m!}\right)\left(\frac{e^{-b\gamma}\gamma^{a-1}b^a}{(a-1)!}\right) d\gamma \\ &= \binom{m+a-1}{m} \left(\frac{\phi(p_n)g(n)}{b+\phi(p_n)g(n)}\right)^m \left(\frac{b}{b+\phi(p_n)g(n)}\right)^a, \\ &\quad \forall m \in \{0,1,2,\ldots\}, \end{aligned} \tag{27}$$

where $f(M_n = m|\Gamma_n = \gamma, p_n)$ is given by (25) and $f(\gamma)$ is the distribution function of Γ_n (i.e., a Gamma distribution with parameters of a and b). From (27), we can see that $f(M_n = m|p_n)$ follows a Negative Binomial distribution with parameters of a and $\frac{b}{b+\phi(p_n)g(n)}$.

Denote m_n as the actual sales in the first period $t = n$. After m_n has been observed, the base demand rate in the next period, Γ_{n-1}, is updated. Using Bayes' rule, the posterior distribution of Γ_{n-1} is as follows:

$$\begin{aligned} f(\gamma|p_n, m_n) &= \frac{f(M_n = m_n|p_n, \Gamma_{n-1} = \gamma) f(\gamma)}{\int_0^\infty f(M_n = m_n|p_n, \Gamma_{n-1} = \gamma) f(\gamma) d\gamma} \\ &= \frac{e^{-\gamma(b+\phi(p_n)g(n))}\gamma^{a+m_n-1}[b+\phi(p_n)g(n)]^{a+m_n}}{(a+m_n-1)!}, \forall \gamma > 0. \end{aligned} \tag{28}$$

Thus, the posterior distribution of Γ_{n-1} at the beginning of the next period ($t = n-1$) follows a Gamma distribution with a scale parameter of $a+m_n$ and a shape parameter of $b + \phi(p_n)g(n)$.

From Bayes' rule, the distribution of demand in period $t = n - 1$ is given by:

$$\begin{aligned}
&f(M_{n-1} = m|p_{n-1}, p_n, m_n) \\
&\quad = \int_0^\infty f(M_{n-1} = m|\Gamma_{n-1} = \gamma, p_{n-1}) f(\gamma|p_n, m_n) d\gamma \\
&\quad = \int_0^\infty \left(\frac{[\phi(p_{n-1})g(n-1)\gamma]^m e^{-\phi(p_{n-1})g(n-1)\gamma}}{m!}\right)\left(\frac{e^{-\gamma[b+\phi(p_n)g(n)]}\gamma^{a+m_n-1}[b+\phi(p_n)g(n)]^{a+m_n}}{(a+m_n-1)!}\right)d\gamma \\
&\quad = \binom{m+a+m_n-1}{m}\left(\frac{\phi(p_{n-1})g(n-1)}{b+\phi(p_n)g(n)+\phi(p_{n-1})g(n-1)}\right)^m \\
&\quad \times \left(\frac{b+\phi(p_n)g(n)}{b+\phi(p_n)g(n)+\phi(p_{n-1})g(n-1)}\right)^{a+m_n}, \forall m \in \{0, 1, 2, \ldots\},
\end{aligned}$$

where $f(M_{n-1} = m|\Gamma_{n-1} = \gamma, p_{n-1})$ is given by (26) for $t = n - 1$, and $f(\gamma|m_n, p_n)$ is the distribution function of Γ_{n-1}, given by (28). We can see that M_{n-1} follows a Negative Binomial distribution with parameters $a + m_n$ and $\frac{b+\phi(p_n)g(n)}{b+\phi(p_n)g(n)+\phi(p_{n-1})g(n-1)}$. From the demand functions described above, we find that the base demand rate and the unconditional ticket demand distributions in any period t can be summarized in the following Theorem.

Theorem 6. *In period t, where $n < t \leq 1$, we have:*

i) the base demand rate, Γ_t, follows a Gamma distribution with a scale parameter of $a + \sum_{k=t+1}^{n} m_k$ and a shape parameter of $b + \sum_{k=t+1}^{n} \phi(p_k)g(k)$,

ii) the unconditional ticket demand, M_t, follows a Negative Binomial distribution with parameters $a + \sum_{k=t+1}^{n} m_k$ and $\frac{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)}{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)+\phi(p_t)g(t)}$.

We present the proofs of Theorem 6 and other key results in the Appendix. Theorem 6 (i) and (ii) address the distribution of the base demand rate, Γ_t, and the unconditional ticket demand, M_t, respectively. From Theorem 6 (ii), the distribution of demand in period t is given by:

$$
\begin{aligned}
&f(M_t = m|p_t, p_{t+1}, \ldots, p_n, m_{t+1}, \ldots, m_n) \\
&\quad = \binom{m + a + \sum_{k=t+1}^{n} m_k - 1}{m} \left(\frac{\phi(p_t)g(t)}{A(t)}\right)^m \left(\frac{b + \sum_{k=t+1}^{n} \phi(p_k)g(k)}{A(t)}\right)^{a + \sum_{k=t+1}^{n} m_k}, \\
&\quad \forall m \in \{0, 1, 2, \ldots\},
\end{aligned}
$$

where $A(t) = b + \sum_{k=t+1}^{n} \phi(p_k)g(k) + \phi(p_t)g(t)$. Note that ticket demand takes non-negative integer values. The posterior Negative Binomial distribution of demand is advantageous since it provides the probability mass function for non-negative integers. Moreover, the shape and position can be changed extensively via its two parameters [135]. From ticket sales data of the 2007-2008 season obtained from a performance venue in the U.S. (108 shows), Figure 30 depicts the probability density plots of sales occurred at different times in the selling horizon; i.e., one, two, three and four months prior to the show time, respectively. Using the Maximum-Likelihood method, the actual data was fitted to the Negative Binomial distribution, where the fitted values of mean and standard deviation were shown. From the Pearsons Chi-Square test [117], we found that with a significant level of 5%, the hypothesis that demand (from the venue's data) follows a Negative Binomial distribution was accepted for all graphs shown in Figure 30.

Given the observed sales data, the seller updates his belief about demand in the next period. The following Proposition presents the relationships between the expected demand and key relevant factors.

Theorem 7. *The expected demand in period t is:*

i) decreasing with the ticket price in period t;

ii) decreasing with selling prices offered in the past periods, $t+1, \ldots, n$;

iii) increasing with sales in the past periods, $t+1, \ldots, n$.

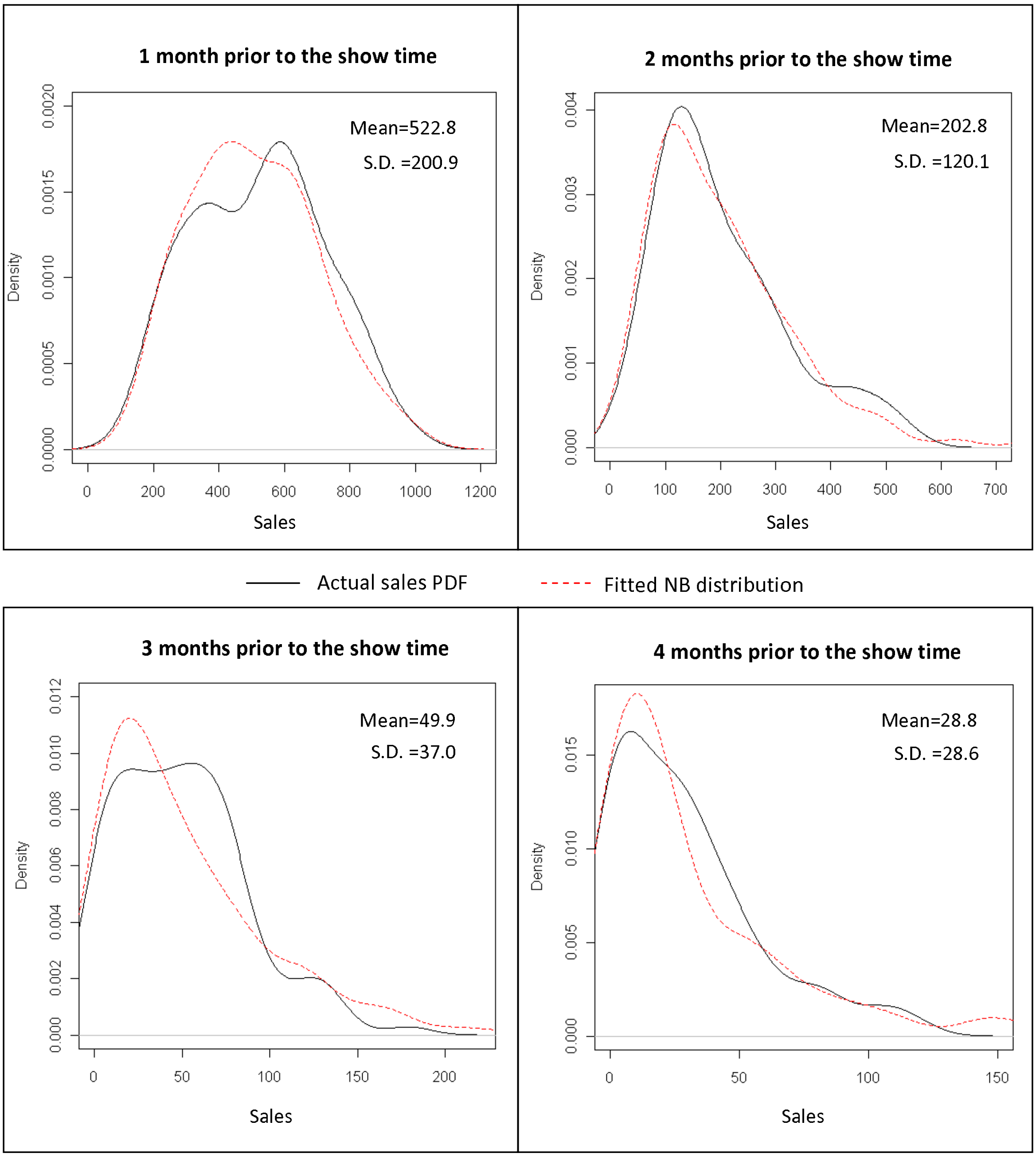

Figure 30: The probability density plots of ticket sales from a major performance venue in the U.S. (during the 2007-2008 season)

From Observation 11, the expected demand in period t depends not only on p_t but also on the history of prices charged, $p_{t+1}, \ldots, p_n$. The result is intuitive because when the seller sold tickets at expensive prices in the past periods, it reduced past sales. This causes lower expected demand in the current period t since the observed sales are used to update belief about the upcoming demand. In other words, the lower (higher) the past sales, the lower (higher) the demand in the upcoming period is expected to be.

So far we have described the demand learning process. Let us next present the dynamic ticket pricing model.

4.4 Dynamic Discounts/Premiums Pricing

In this section we describe how the demand learning technique discussed in Section 4.3 can be embedded in the ticket seller's dynamic pricing model. We begin with descriptions of key assumptions and present the "Dynamic discounts/premiums pricing" model, where price discounts/premiums adjustments are allowed in every period, in the context of a stochastic dynamic program in Section 4.4.1. Computational analysis is then presented in Section 4.4.2 to provide insights on the optimal pricing structures and the benefit of demand learning for the described pricing model.

4.4.1 Model

We consider a discrete finite time setting, where there are n sub-periods from the beginning of the selling horizon to the event/show time. The time periods are numbered in reverse chronological order so that the beginning of the selling horizon is time $t = n$, and the show takes place at time $t = 0$. At the beginning of the selling horizon ($t = n$), there are I_n number of tickets for sale. The number of tickets demand and the actual sales in period t are denoted by d_t and m_t, respectively. If demand is less than inventory at the beginning of period, I_t, the ticket sales in period t (m_t) are equal to d_t. However, if demand exceeds inventory ($d_t > I_t$), the ticket sales m_t will equal I_t.

In the first period of the selling horizon ($t = n$), the seller determines the ticket base price, p_n. Then, at time $t = n - 1, \ldots, 1$, the seller observes past sales, and decides whether to offer any discounts (reduce price) or charge premiums (increase price), and how much. Denote θ_t as the discount or premium for period t, and we assume it is based on p_n (i.e., customers pay $\theta_t p_n$ for each ticket in period t). If $\theta_t < 1$, then θ_t is considered as a price discount; otherwise, it is a price premium. Define a_t as the ticket seller's action (or decision) at time t; we have:

$$a_t = \begin{cases} p_n & \text{if } t = n \\ \theta_t & \text{if } t = n-1, \ldots, 1. \end{cases}$$

Let $V_t(h_t)$ denote the maximum expected revenue obtained from time period t until the time of the show/event (at $t = 0$), given the history at the beginning of period t is $h_t = (h_{t+1}, a_{t+1}, m_{t+1}, I_t)$. Note that h_t consists of the history of the previous period $t+1$ (i.e., the action and the number of ticket sold in the previous period $t+1$), as well as the current inventory level (I_t). At the beginning of the selling horizon ($t = n$), we have $h_n = I_n$ since there is no prior information. We assume the seller is risk neutral, so the backwards recursion of the revenue maximization problem can be written as:

$$V_t(h_t) = \max_{a_t \in A_t} E[R_t(a_t) + V_{t-1}(h_{t-1}) | m_n, \ldots, m_{t+1}; \phi(p_n), \ldots, \phi(p_{t+1}); g(n), \ldots, g(t+1)], \tag{29}$$

where $R_t(a_t)$ is defined as the immediate revenue received in period t when the seller's action is a_t; i.e.,

$$R_t(a_t) = \begin{cases} a_n \min\{I_n, d_n(a_n)\} = p_n \min\{I_n, d_n(p_n)\} & \text{if } t = n; \\ a_t p_n \min\{I_t, d_t(a_t)\} = \theta_t p_n \min\{I_t, d_t(\theta_t p_n)\} & \text{if } t = n-1, \ldots, 1, \end{cases}$$

where $I_t = \max\{0, I_{t+1} - d_{t+1}(a_{t+1})\}$. The boundary conditions are as follow:

$$V_0(h_0) = 0, \forall h_0, \tag{30}$$

$$V_t(h_t = (h_{t+1}, a_{t+1}, m_{t+1}, 0)) = 0, \forall t. \tag{31}$$

Condition (30) means there is zero salvage value of any unsold tickets at the event time ($t = 0$). Condition (31) states that when all tickets have already been sold, the revenue from any period t on is zero since there are no tickets left for sale. The stochastic dynamic program can be solved by backwards induction starting at the final period, $t = 1$, to period $t = n$. The following example is a special case where the selling time is divided into two periods, $n = 2$.

Example: Special case when n=2

The first period of the selling season is at $t = 2$, where the ticket seller determines the optimal base price p_2. Then, after observing sales in period $t = 2$, m_2, the seller determines the price discount/premium for the next period $t = 1$. This problem can be solved by backwards recursion, and the revenue maximization problem at $t = 1$ can be written as

$$\begin{aligned} V_1(I_2, p_2, m_2, I_1) &= \max_{\theta_1 \geq 0} E[R_1(\theta_1)|m_2, \phi(p_2), g(2)] \\ &= \max_{\theta_1 \geq 0} E[\theta_1 p_2 \min\{I_1, d_1(\theta_1 p_2)\}|m_2, \phi(p_2), g(2)], \end{aligned}$$

where $I_1 = \max\{0, I_2 - d_2(p_2)\}$. Given the inventory, price, and sale in the last period are I_2, p_2, m_2, respectively, the unconditional distribution of demand $(d_1(\theta_1 p_2))$ in the current period $t = 1$ when offering a price discount/premium (θ_1) is Negative Binomial with parameters $a + m_2$ and $\frac{b+\phi(p_2)g(2)}{b+\phi(p_2)g(2)+\phi(\theta_1 p_2)g(1)}$.

At the beginning of the selling period ($t = 2$), the seller solves the following optimization problem:

$$\begin{aligned} V_2(I_2) &= \max_{p_2 \geq 0} E[R_2(p_2) + V_1(I_2, p_2, m_2, I_1)] \\ &= \max_{p_2 \geq 0} E[p_2 \min\{I_2, d_2(p_2)\} + V_1(I_2, p_2, \min\{I_2, d_2(p_2)\}, \big(I_2 - d_2(p_2)\big)^+)], \end{aligned}$$

since $m_2 = \min\{I_2, d_2(p_2)\}$ and $I_1 \max\{0, I_2 - d_2(p_2)\} = \big(I_2 - d_2(p_2)\big)^+$. Note that in this period, the distribution of demand $(d_2(p_2))$ when offering the base ticket price, p_2, is Negative Binomial with parameters a and $\frac{b}{b+\phi(p_2)g(2)}$, respectively. To observe the characteristics of the optimal solution and the benefits from the learning process under different scenarios, computational experiments are conducted in the next section.

4.4.2 Computational Experiments

In this section, we perform computational experiments for the dynamic discounts/premiums pricing model, with the goal of understanding the properties of the optimal solutions and the model performance under different situations. Specifically, we consider: (1) how the remaining inventory affects the expected base demand rate and the optimal discounts/premiums pricing policy, and (2) how the performance of the dynamic discounts/premiums pricing model with demand learning is compared to the model without demand learning and when demand learning is most beneficial.

4.4.2.1 Effects of the Remaining Inventory of Unsold Tickets

In each period, the seller can observe how many unsold tickets remained in inventory. This section explores how the seller's expectation and optimal decisions are affected by the inventory information. Let us consider the following example. A ticket seller has 300 tickets for sale in a selling season with $n = 2$. Let the probability of each arrival purchasing a ticket (the price effect function) be $\phi(p_t) = e^{-wp_t}$, where w is a known non-negative scalar. Note that with this form of price effect, the demand rate equals $g(t)\Gamma_t$ when p_t equals 0 and the rate approaches 0 as p_t is very large. This form of exponential price effect function has been generally used in the marketing literature since it was found to very well fit with the empirical data [1].

Let the timing effect be $g(2) = 1$ and $g(1) = u$, (where $u > 1$ since we assume higher demand in the period closer to the show date). At the beginning of the selling horizon ($t = 2$), the seller believes that ticket demand is Poisson, with the base rate (Γ_2) following a Gamma distribution with parameters $a = 4$ and $b = 0.04$[2].In this period ($t = 2$), the seller determines the optimal base ticket price, p_2^*, from a discrete set P_2=$\{30, 35, \ldots, 75, 80\}$. After he observes how many tickets have been sold, the seller updates his belief with demand learning techniques presented in Section 4.3,

[2]This choice of a and b follows from an example of actual ticket demand data we observed.

and determines the optimal price discount/premium, θ_1^*, to charge in period $t = 1$ from a discrete set $\theta_1=\{70\%, 75\%, \ldots, 115\%, 120\%\}$.

Figure 31 presents the updated mean demand for period $t = 1$ at different values of remaining inventory from the past period (where the price sensitivity parameters (w) equal 0.018, 0.020, and 0.022, respectively[3]). As expected, at the same values of remaining inventory, the graph for $w = 0.018$ is the highest and the graph for $w = 0.022$ is the lowest in Figure 31, since a higher price sensitivity leads to lower expected demand. Moreover, considering the trends, we can see that the higher number of tickets left in inventory (i.e., the lower number of tickets already sold), the lower mean demand the seller expects for the next period. This is consistent with analytical results in Theorem 7 (iii) in Section 4.3 that the expected mean demand is increasing with observed sales in the past periods (so it is decreasing with the leftover inventory). The finding is rational because having a large number of tickets sold in the past periods (or a few number of tickets left in the inventory) can be an indicator of show popularity and it is likely that high demand will occur in the next period as well.

We solve the described problem by the stochastic dynamic program described in Section 4.4.1. The optimal base price at the beginning of the selling horizon ($t = 2$) is 95. After sales are realized, the seller's optimal discount/premium pricing for period $t = 1$ is shown in Figure 32, where the horizontal axis represents different values of leftover inventory, I_1. Note that different lines in Figure 32 (a) represent different price sensitivity parameters, i.e., w equals 0.018, 0.020, and 0.022, respectively; while different lines in Figure 32 (b) represent different timing effect parameters, i.e., u equals 1.7, 2.0, 2.3, respectively.

Observation 11. *The optimal ticket price discount/premium, θ_t, is non-increasing*

[3]Parameter values for the price sensitivity parameters (w) in this experiment result in shifts of 0.9%-1.1% (at p_1=50) which is consistent with data we have seen.

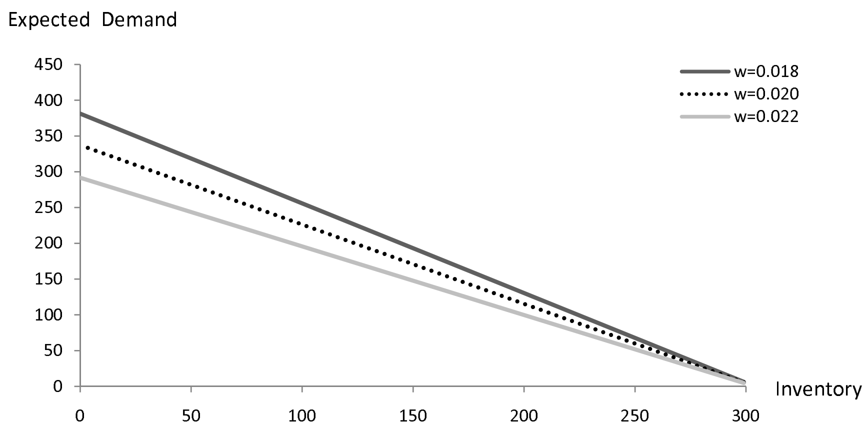

Figure 31: The expected demand for the last period ($t = 1$) at different values of remaining inventory

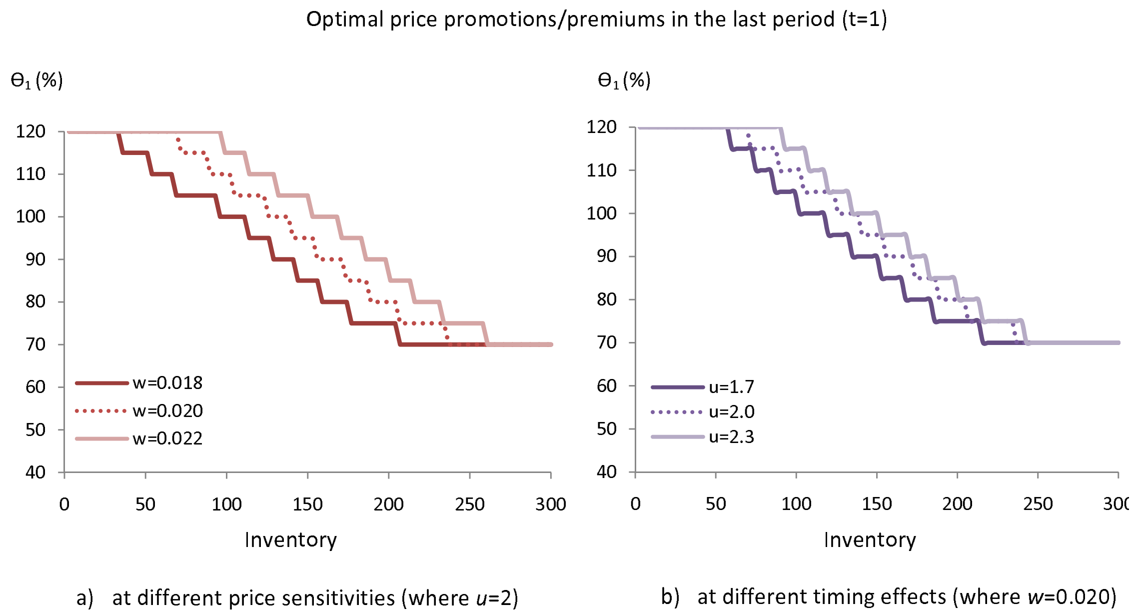

Figure 32: The optimal discounts/premiums pricing for the last period ($t = 1$) at different values of leftover inventory, where: (a) w=0.018, 0.020, and 0.022; (b) u=1.7, 2.0 and 2.3

with the amount of leftover inventory observed at the beginning of the period.

Similar to other sets of experiment results we have explored, Observation 11 states that the higher the inventory, the lower the percentage of the base price that should be charged in the last period ($t = 1$). This result is consistent with other studies of dynamic pricing with no demand learning where the optimal price is found to be non-increasing with the remaining inventory (see [30], [49] and [149]). In addition, we can see in Figure 31 (a) that at the same level of inventory, the higher the price sensitivity, the lower the optimal θ_1. The timing effect parameter (u) also impacts the optimal θ_1. From Figure 31 (b), at the same inventory level, the line with $u = 2.3$ is the highest and the line with $u = 1.7$ is the lowest It implies when high demand is expected in the last period, it is less necessary for the seller to offer a deep discount (since there is a lower risk of having empty seats).

Next, we examine benefits of demand learning and identify which situations are most worthwhile for applying the learning process.

4.4.2.2 Benefits of Demand Learning Process

In this section, we study how the performance of the dynamic discount/premium pricing model with demand learning compares to the model without demand learning. The insights will allow us to identify which situations are most worthwhile to apply the learning process. The demand learning (DL) model incorporates observed sales in updating the belief of the next period's demand, while the no demand learning (NoDL) model does not. Thus, when the seller applies the NoDL model, in every period he uses the same estimated base demand rate as in the first period. We capture the DL and the NoDL model performances by comparing revenues obtained from each model to the revenue under perfect information. Note that under perfect information, the seller knows the true base demand without uncertainty and he optimizes his prices based on the true value. Define PFM_{DL} as the performance of demand learning, i.e., $PFM_{DL} = \frac{\texttt{Revenue of demand learning model}}{\texttt{Revenue of perfect information case}} \times 100$, and PFM_{NoDL} as the performance of the no demand learning model, i.e., $PFM_{NoDL} = \frac{\texttt{Revenue of no demand learning model}}{\texttt{Revenue of perfect information case}} \times 100$. The closer PFM_{DL} and PFM_{NoDL} are to 100, the better their performances are as compared to the perfect information case, although the latter may not be achievable.

Consider a ticket seller having 100 tickets for sale in a two-period selling horizon, where the timing effects are $g(2) = 1$ and $g(1) = u$, $u > 1$. The price effect function is $\phi(p_t) = e^{-0.02p_t}$. At the beginning of the selling time ($t = 2$), the seller determines the optimal base ticket price, p_2^*, from a discrete set P_2=$\{50, 55, \ldots, 95, 100\}$. Then in the next period ($t = 1$), he determines the optimal price discount/premium (θ_1^*) from a discrete set Θ_1=$\{70\%, 75\%, \ldots, 115\%, 120\%\}$.

Let us consider what happens if the ticket seller misestimates demand. For instance, suppose the true base demand rate in the first period ($t = 2$) is γ, while the seller believes the base demand rate follows Gamma distribution with parameters a and b, respectively. Note that the expected value of the base demand rate in this case equals $\frac{a}{b}$ and the variance equals $\frac{a}{b^2}$ (for Gamma distribution). Thus, the demand is overestimated if the seller believes that $\frac{a}{b}$ is greater than the true base demand rate, γ, and the demand is underestimated otherwise.

In the following example, let the true base demand rate, γ, be 120. Figure 33 shows the chosen prices for the first and the second periods when the seller's estimates of the mean base demand rate ($\frac{a}{b}$) are 30, 60, ..., 210, 240, respectively. The second period ($t = 1$) price equals the first period ($t = 2$) price times the discount/premium of the second period ($p_1 = p_2 \times \theta_1$). In case 1, u=2.0, and the optimal prices under perfect information are 75 and 56 for the first and the second period (in case 2 with u=2.3, they are 80 and 60, respectively). We can see from Figure 33 that in the first period, both DL and NoDL models choose similar prices for almost all values of $\frac{a}{b}$. However, when the base demand rate is underestimated, i.e., $\frac{a}{b} < 120$, the DL model's prices for the second period are higher than the NoDL model's. Due to the ability of updating demand, the DL model uses the observed sales to adjust the true base demand rate's distribution for the next period. With the DL model, the seller realizes in the second period that he has underestimated the rate and then decides to charge higher prices than he would have done without demand learning (NoDL). Similar reason applies when the base demand rate is overestimated, where the second period's prices chosen by the DL model are lower than the NoDL model.

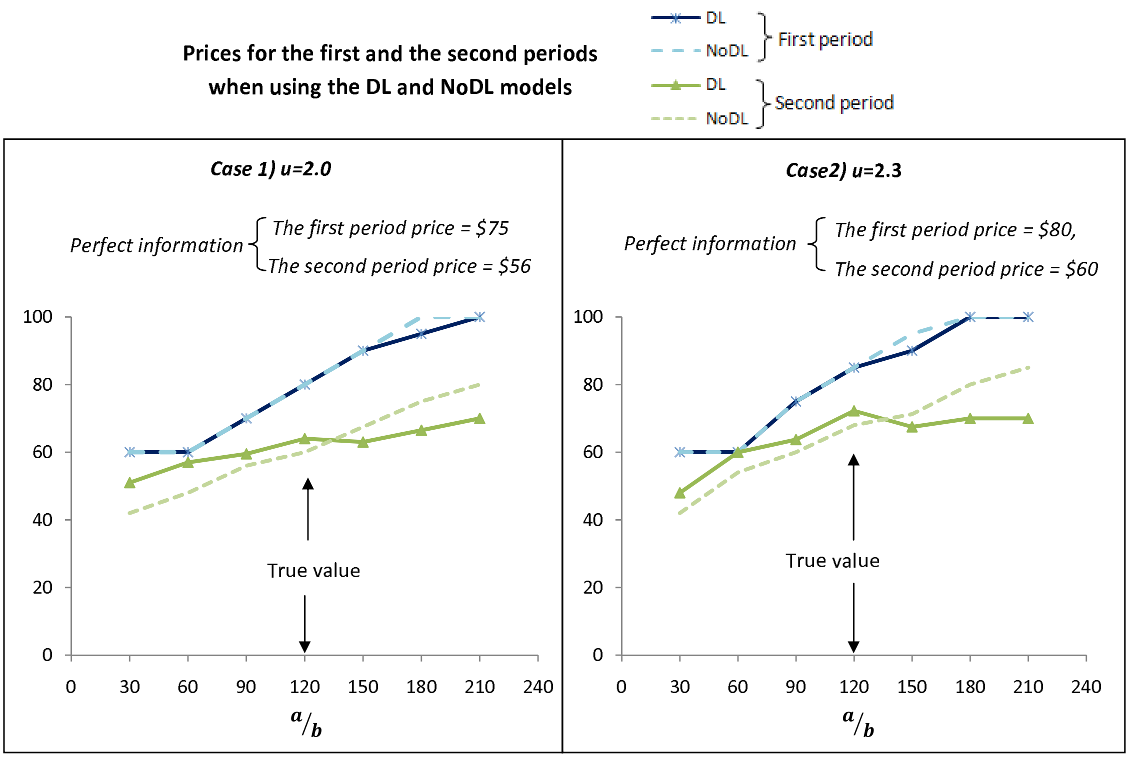

Figure 33: The optimal prices for the first and the second periods when using the DL and the NoDL models

Considering the performance of each model, Figure 34 shows PFM_{DL} and PFM_{NoDL}, where the horizontal axis represents different estimates of $\frac{a}{b}$ and the true value is indicated.

Observation 12. *We found that:*

i) Demand learning (DL) is most beneficial, as compared to no demand learning (NoDL) when the initial estimation is inaccurate (with up to 8-11% improvement in revenue when the misestimates are high).

ii The marginal benefit of the DL model over the NoDL model is higher when

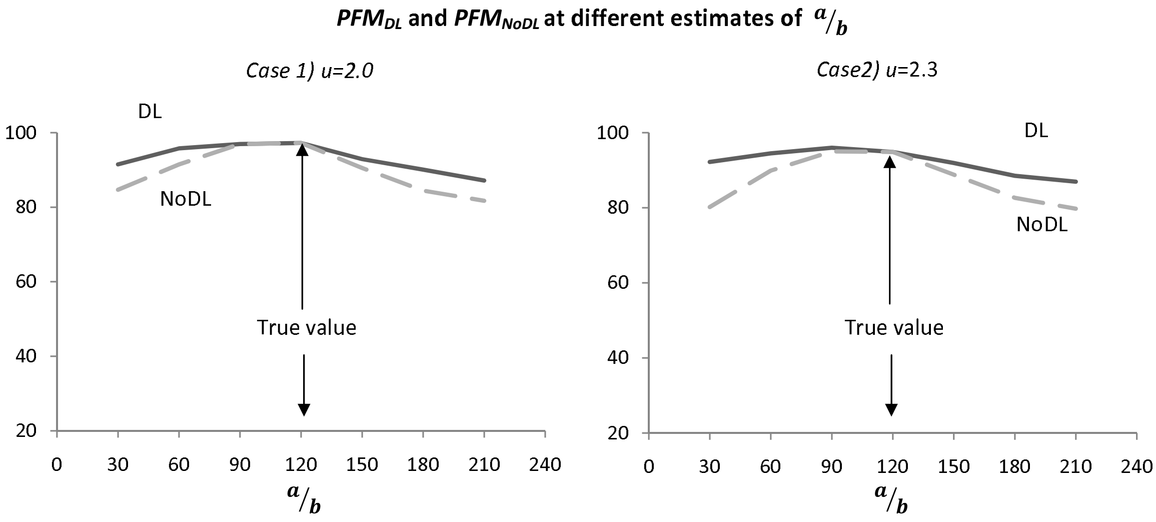

Figure 34: Performances of the DL and the NoDL models at different estimates of $\frac{a}{b}$, when the timing effects, u, are 2.0 (the left diagram) or 2.3 (the right diagram)

a greater amount of demand arrives in the last period (approximately 5.4% improvement in revenue on average for case 2, as compare to 3.9% for case 1).

iii) The underestimation of the base demand rate when using the DL model causes fewer revenue loss as compared to overestimation.

As we can see in Figure 34, when the misestimates of the base demand rate are fairly large, PFM_{DL} is greater than PFM_{NoDL} for both graphs. It implies demand learning can help increase revenue, compared to no learning. The intuition is that when there are estimation errors, observed sales are effectively used since the inaccurate estimation can be corrected for an updated belief of demand in the next period. For example, if there is high uncertainty about the show's popularity (e.g., a new show with no records of past sales), it can be difficult for the ticket seller to identify a correct value of the show's base demand rate. The seller can start with a rough estimate of the base demand rate and then use the observed sales to update his belief.

Comparing case (1) and case (2) in Figure 33, we found the marginal improvement of the DL model over the NoDL model is larger when a greater amount of demand arrives later in the last period (where the percentage improvements are 3.91% and 5.44% for case (1) and (2), respectively). An explanation is that when a large amount of demand arrives in the last period, it can be more critical to optimally choose the price for that period since wrong prices can lead to a greater loss of the potential revenue. The demand learning model has a greater advantage over the No demand learning model because it uses observed sales to update the belief about demand, and thus it can be more effective in choosing an optimal price for the last period.

As we observe in Figure 33, when the base rate is overestimated, the seller chooses to offer higher prices than the true optimal, leading to a drop in ticket demand and the overall revenue. On the other hand, when the ticket seller underestimates the base demand rate, the models suggest lower prices. The resulting demand is then higher than the optimal while the revenue is less. From Figure 34, when the seller uses observed sales to update his belief, underestimation of the base demand rate leads to less revenue loss than overestimation (Observation 12 (iii)). It implies a risk-averse ticket seller who tends to underestimate the demand rate may receive greater revenue than a risk-taker seller who overestimates it. In other words, underestimation of demand is fairly safer than overestimation. We note that these results are consistent with other sets of experiments (with different model parameters) we have performed.

So far, we have considered the dynamic discount/premium pricing where the price adjustment is allowed every period, prior to the show date. In the next section, we consider an alternative practice in which the offering time of discounts/premiums is also a decision variable and there is only one opportunity to adjust price in the selling horizon.

4.5 Dynamic Timing for Price Discounts/Premiums

In practice, changing price too often may not be preferable for some performance organizations who believe it may reduce customer satisfactions. From our discussions with the Sports and Entertainment ticket sellers, an interesting question they point out is: if they will change ticket price only once in the selling period by reducing or increasing 5% or 10% from the original price, when should they do so and by how much (based on demand they have seen so far)? In this section, we consider an alternative model that dynamically determines the optimal time and level for price discounts/premiums during the selling season. After the seller has already offered a discount or charged a price premium, the price will remain the same in all following periods.

4.5.1 Model

We note that the notation of parameters and variables in this section is similar to what we have defined in Section 4.4.1. When there is only one chance to adjust price, the backwards recursion of the revenue maximization problem is given by:

$$V_n(I_n) = \max_{p_n \in P_n} E[p_n \min\{I_n, d_n(p_n)\} + V_{n-1}(h_{n-1})],$$

for the first period ($t = n$), where P_n is the set of allowable prices and

$$\begin{aligned} V_t(h_t) \quad &= \quad \max\Big\{\max_{\theta_t \in \Theta_t} E[\theta_t p_n \sum_{i=1}^{t} \min\{I_i, d_i(\theta_t p_n)\} | B_t], \\ &\qquad E[p_n \min\{I_t, d_t(p_n)\} + V_{t-1}(h_{t-1}) | B_t]\Big\}, \forall t = n-1, \ldots, 1, \end{aligned}$$

where $B_t = m_n, \ldots, m_{t+1}; \phi(p_n), \ldots, \phi(p_{t+1}); g(n), \ldots, g(t+1)$, and Θ_t is the set of discrete allowable levels of discount/premium. A discrete list of prices is also considered in the dynamic pricing with no demand learning literature (e.g., [30], [49] and [48]) since it is easy to be implemented and controled. Note that $I_t = \max\{0, I_{t+1} - d_{t+1}(\theta_{t+1} p_n)\}$ if the seller has decided to offer a price discount/premium, and $I_t = \max\{0, I_{t+1} - d_{t+1}(p_n)\}$ otherwise.

In the first period ($t = n$), the seller chooses the base price p_n from the set P_n. Then, for each of the future periods, the seller evaluates the expected revenue from: 1) offering a price premium/discount, starting from this period, or 2) delaying the price discount/premium decision to one of the future periods by continuing with the base price p_n in this period. For the first choice, the expected revenue is $E[\theta_t^* p_n \sum_{i=1}^{t} \min\{I_i, d_i(\theta_t^* p_n)\}|B_t]$ since the final price is $\theta_t^* p_n$ for all the rest of the periods and there is no further decisions. For the second choice, the expected immediate plus future revenue equals $E[p_n \min\{I_t, d_t(p_n)\} + V_{t-1}(h_{t-1})|B_t]$ since the seller's price is still p_n and the decision repeats in the next period.

The boundary conditions are similar to equations (30) and (31) in section 4.4.1, where $V_0(h_0) = 0, \forall h_0$ and $V_t(h_t = (h_{t+1}, a_{t+1}, m_{t+1}, 0)) = 0, \forall t$. The stochastic dynamic program above can be solved by backwards recursion from the final period, $t = 1$, to period $t = n$. We present key computational results for managerial insights in the next section.

4.5.2 Computational Experiments

In this section, computational experiments are performed to further analyze the solutions of the dynamic timing for price discounts/premiums. The results can give insights on when the seller should offer a discount/premium early in the selling horizon, or close to the show date. We also compare the dynamic timing model's performance to the original dynamic discount/premium pricing model to identify the impact of restricting the number of price adjustments.

4.5.2.1 Solutions of the Dynamic Timing Model

Consider a selling season with three periods, $n = 3$. In the first period, $t = 3$, the seller determines the base price p_3 from a discrete set P_3=$\{50, 55, \ldots, 95, 100\}$. After learning demand from the observed sales, in period $t = 2$ he decides if any discounts/premiums (chosen from the set $\{70\%, 75\%, \ldots, 115\%, 120\%\}$) should be used from this period on, or he should continue with the base price and repeat this decision again in a future period (at $t = 1$). Figure 35 presents the optimal price discounts/premiums for period $t = 2$ under different values of the remaining inventory.

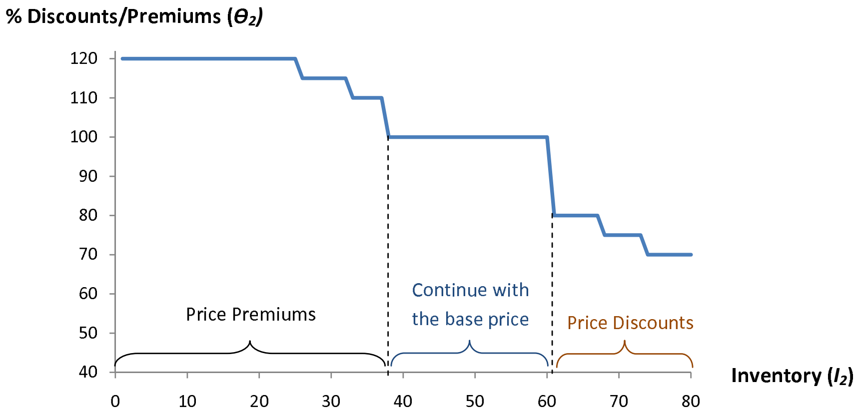

Figure 35: The optimal price discounts/premiums for period $t = 2$ at different values of remaining inventory (I_2), where the starting inventory, $I_3 = 80$

Observation 13. *The optimal decisions for period* $t = 2$ *are as follow.*

i) Offer price discounts ($\theta_2^* < 1$) *when the remaining inventory is high. The optimal percentage of the base price decrease* ($1 - \theta_2^*$) *is non-decreasing with the level of unsold inventory.*

ii) Charge price premiums ($\theta_2^* > 1$) *when the remaining inventory is low. The optimal percentage of the base price increase* ($\theta_2^* - 1$) *is non-increasing with the level of unsold inventory.*

ii) Continue with the base price, p_3^**, and delay the decision of price discount/premium to the next period when the remaining inventory is moderate.*

From Observation 13, it is not always optimal to delay the pricing decisions in an effort to learn more about demand. We can see in Figure 35 that the seller should use discounts/premiums in period $t = 2$ when either low or high demand was observed in the previous period (equivalently, either high or low remaining inventory). The explanation is that when the observed sales are low (i.e., there are a large number of unsold tickets), the future demand is also expected to be low. It is optimal to use price discounts to induce more revenue in this period, rather than delaying the decision. Similarly, when the observed sales are high (or higher than the seller's expectation), it can imply the seller's belief of the base demand rate should be increased and/or the decided base price is too low. Thus, the model suggests increasing price with price premiums, starting from this period. However, for a moderate level of remaining inventory, it can imply the base price is already appropriate and it is unnecessary to use discounts/premiums in this period. Moreover, it can mean that more information is necessary for price changing decisions. So, for those moderate levels of inventory, the seller should continue with the current price and delay the decision to the future period. These observations are consistent with other sets of numerical experiments we have performed, using different model parameters.

As we have described, if the seller decides to use a discount or premium in period $t = 2$, the selected discount/premium will be applied in the last period ($t = 1$), and thus the net ticket price remains the same until the show date. Let us next consider the optimal price discount/premium in the last period, $t = 1$, given that a discount/premium has not been used previously. Figure 36 depicts the optimal last-period discounts/premiums at different levels of remaining inventory, I_1, where the inventory in the previous period equals I_2. We observe that for all values of I_2, the optimal ticket price discount/premium for the last period (θ_1) is non-increasing with the amount of remaining inventory observed at the beginning of this period, I_1. This result is consistent with what we have seen in Observation 11 of the original discounts/premiums pricing model in Section 4.4.

When the number of price changes is limited to be at most only once, we have seen that different levels of remaining inventory can lead to different decisions, i.e., either offering discounts/premiums immediately or delaying decisions to the next period. It is interesting to identify the dynamic timing model's performance as compared to the

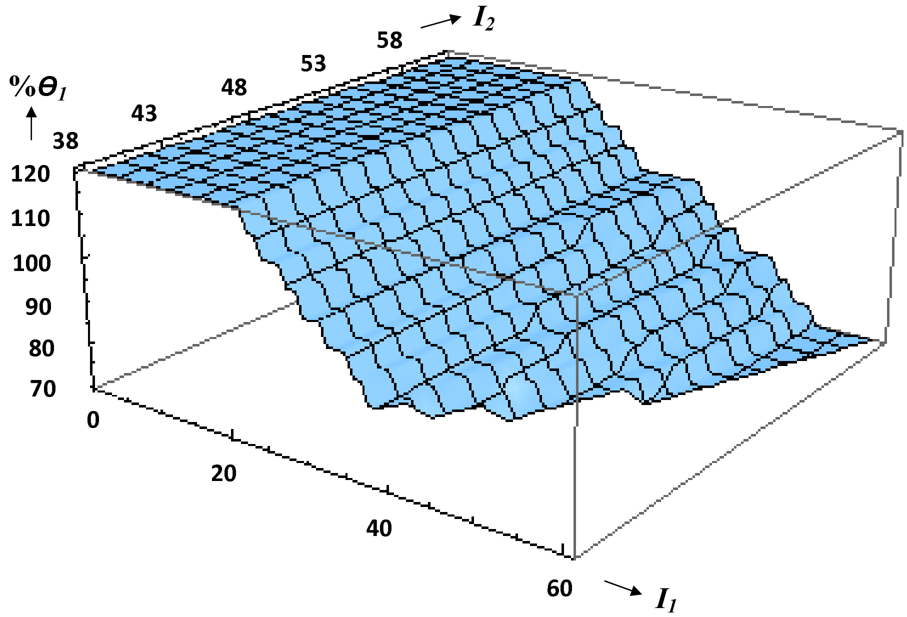

Figure 36: The optimal discounts/premiums for the last period ($t = 1$) at different levels of remaining inventory (I_1), given that the inventory at time $t = 2$ equals I_2

original dynamic discounts/premiums pricing model. We generate some insights on this research question in the following analysis.

4.5.2.2 Comparisons between the Dynamic Timing Model and the Original Dynamic Discounts/Premiums Pricing Model

In this section, we examine the performance of the dynamic timing model, as compared to the original discounts/premiums pricing model. In some situations, it may be unnecessary to offer different discounts/premiums every period, while in other cases it can be useful to do so. Our goal is to obtain insights on how the seller should effectively select models, studied in Section 4.4.1 and 4.5.1, for different circumstances. Denote PFM_{DT} as the performance of the dynamic timing model (compared to the dynamic discounts/premiums pricing model), where $PFM_{DT} = \frac{\texttt{Dynamic timing model's revenue}}{\texttt{Dynamic discounts/premiums pricing model's revenue}} \times 100$. When PFM_{DT} is close to 100, the timing model's performance approaches the dynamic discounts/premiums pricing model's.

Consider a seller having 300 tickets to sell in a horizon with four periods, $n = 4$, where the base price (p_4) is selected from a discrete set $\{50, 55, \ldots, 95, 100\}$ and the discounts/premiums are chosen from the set $\{70\%, 75\%, \ldots, 115\%, 120\%\}$. We examine the impact of demand variation by measuring PFM_{DT} when the base demand rate (Γ) follows a Gamma distribution with mean $\frac{a}{b} = 100$, 130 or 160. For each value of the mean base demand rate, we consider different parameters a and b for different values of standard deviation ($\frac{\sqrt{a}}{b}$), presented in the table in Figure 37. Note that the graph in the bottom of Figure 37 shows the dynamic timing model's performance (compared with the original discounts/premium pricing model), PFM_{DT}, at different values of base demand rate's standard deviation[4].

Observation 14. *It is less (more) necessary for the ticket seller to vary price discounts/premiums every period when the base demand rate's standard deviation is low (high).*

From Figure 37 , we observe that as the standard deviation of the base demand rate (Γ) increases, PFM_{DT} decreases, implying the dynamic timing model's performance is further from the original discounts/premiums pricing model's. In other words, if the seller expects high variation of the base demand rate, it can be useful to often adjust the discounts/premiums to effectively capture the change of demand. This is because when the the standard deviation of the initial estimate is high, the seller cannot rely much on the initial estimate since there is a high probability that it takes on other values. The seller may need to frequently adjust the discounts/premiums as there is a variety of base demand rates in each period. On the other hand, limiting a price change to only once is suitable for low demand variation. In this case, the same price can be appropriate for many consecutive periods since the demand rate is less likely to be much higher or lower than what is expected. This observation addresses

[4]The price effect and the timing effect functions for Figure 37 are $\phi(p_t)$ when $t = 4$, $\phi(\theta_t p_4) = e^{-0.025\theta_t p_4}$ when $t = 1, \ldots, 3$, and $g(t) = 5 - t, \forall t = 1, \ldots, 4$.

i) a/b = 100	a	90	70	50	40	30	20
	b	0.90	0.70	0.50	0.40	0.30	0.20
	S.D.	**10.54**	**11.95**	**14.14**	**15.81**	**18.26**	**22.36**
	PFM_{DT}	**98.04**	**97.80**	**97.36**	**96.89**	**96.40**	**95.71**
ii) a/b = 130	a	150	120	100	80	60	35
	b	1.15	0.92	0.77	0.62	0.46	0.27
	S.D.	**10.61**	**11.87**	**13.00**	**14.53**	**16.78**	**21.97**
	PFM_{DT}	**97.85**	**97.63**	**97.35**	**97.17**	**96.69**	**96.26**
iii) a/b = 160	a	250	210	170	130	90	50
	b	1.56	1.31	1.06	0.81	0.56	0.31
	S.D.	**10.12**	**11.04**	**12.27**	**14.03**	**16.87**	**22.63**
	PFM_{DT}	**97.53**	**97.41**	**97.31**	**97.21**	**97.01**	**96.46**

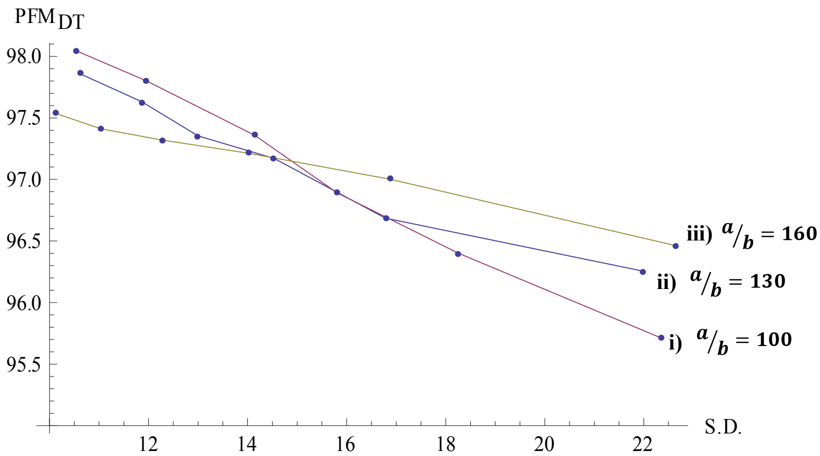

Figure 37: Performance of the dynamic timing model, PFM_{DT}, at different values of the base demand rate's standard deviation, where $\frac{a}{b}$=100, 130 and 160, respectively

that the seller should consider the base demand rate's variation when deciding how often the ticket price should be adjusted.

In addition to demand variation, we consider the impact of the timing effects on the dynamic timing model's performance. Depending on the show's popularity, venue locations, etc., the pace of demand in each selling period can be different. For instance, much of the popular performance's ticket demand may arrive early in the selling season since people expect those tickets to be sold out soon, while most demand of the less popular shows may arrive later or closer to the show date. In the next analysis, we explore whether the seller should limit the frequency of price changes by using the dynamic timing model, or whether he should adjust price every period by using the original discounts/premiums pricing model, for shows having different demand paces.

We consider the following timing-effect functions for a selling horizon with 4 periods;

A) $g(4)=1$, $g(3)=3$, $g(2)=5$, $g(1)=7$,

B) $g(4)=1$, $g(3)=2.6$, $g(2)=4.6$, $g(1)=7.8$,

C) $g(4)=1$, $g(3)=2.2$, $g(2)=4.2$, $g(1)=8.6$,

D) $g(4)=1$, $g(3)=1.8$, $g(2)=3.8$, $g(1)=9.4$,

E) $g(4)=1$, $g(3)=1.4$, $g(2)=3.4$, $g(1)=10.2$,

F) $g(4)=1$, $g(3)=1$, $g(2)=3$, $g(1)=11$,

where the timing-effect trend for each case is presented in Figure 38. Note that the average of the timing effects for all cases ($\sum_{i=1}^{4} g(t)/4$) is equal to 4, to be consistent in average demand. In case A, demand is expected to increase linearly as it is closer to the show date; while the ticket demand in case F does not increase much at the beginning and most of the demand arrives when it is close to the show date.

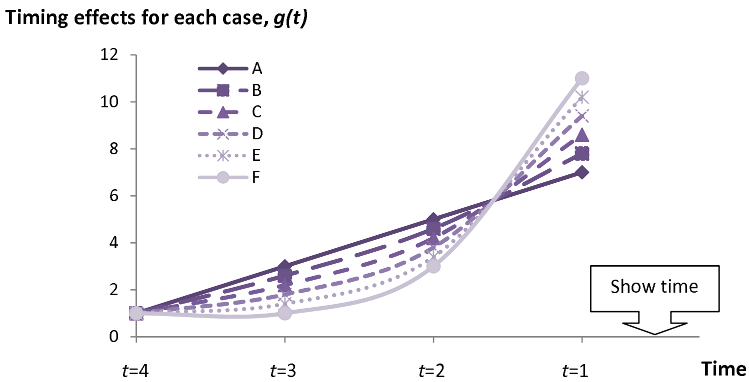

Figure 38: The timing effects for cases A, B, . . . , F, respectively

To capture how the dynamic timing model performs as compared to the original dynamic discounts/premiums pricing model, PFM_{DT} of each case is measured. In addition, we consider two conditions: (1) demand learning is allowed, and (2) demand learning is not allowed, because it is interesting to see if demand learning affects the model's performances. Figure 39 shows PFM_{DT} for cases A, B, ..., F, respectively, under (1) demand learning and (2) no demand learning [5].

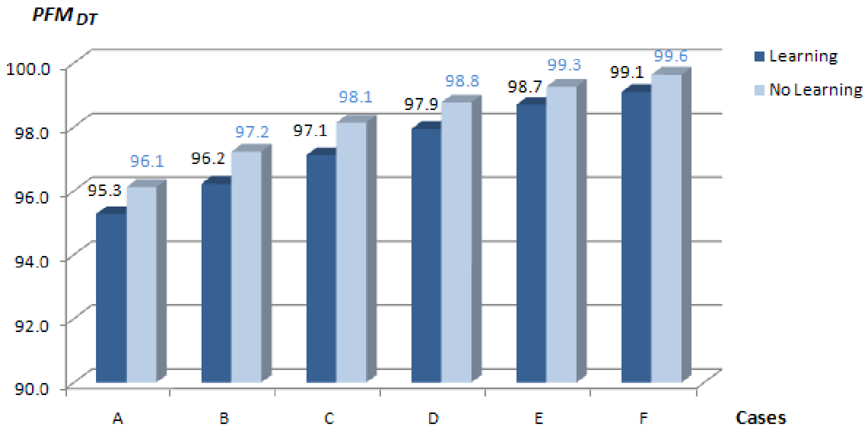

Figure 39: The performance of the dynamic timing model for each case (as compared to the discounts/premiums pricing model), when there are (1) demand learning, and (2) no demand learning

From Figure 39, PFM_{DT} is lowest in case A, implying that the performance of limiting the number of price changes (as compared to allowing price adjustment every periods) is lowest when the timing effect increases linearly. As a larger amount of demand arrives closer to the end of the selling horizon (from case B to case F), PFM_{DT} increases, showing the dynamic timing model's performance is closer to the original discounts/premiums pricing model's revenue.

In addition to the demand pace, we observe that demand learning affects the performance difference between the two models. Specifically, *the dynamic timing model's*

[5] Parameter values are $a = 100$, $b = 0.77$, $I_4 = 300$, and the price effect function, $\phi(p_t) = \phi(\theta_t p_4) = e^{-0.025\theta_t p_4}$.

performance is closer to the dynamic discounts/premiums pricing model's when demand learning is not allowed (as PFM_{DT} is closer to 100 in this case). It indicates limiting the frequency of price changes is less disadvantageous, compared to allowing price changes every period, when the seller does not use observed sales to update his belief of demand. An explanation is when the seller allows demand learning, new information is received every period. So, it can be more beneficial to change price often, according to the updated demand. Adjusting price regularly as in the dynamic disconts/premiums pricing model can lead to higher revenue, especially with an integration of demand learning implementation. This observation is consistent with results across the set of computational experiments we performed, using different demand parameters.

We conclude this section by exploring the benefit gained from dynamic pricing as compared to static pricing. Figure 40 shows the average percentage revenue increase from implementing the dynamic discounts/premiums pricing model and the dynamic timing model, with demand leaning and without demand learning, respectively. We observe that dynamic pricing (without demand learning) can help increase revenue by approximately 3.26-4.33%, compared to static pricing. When demand learning is incorporated with one price change (dynamic timing model), the average revenue increase is approximately 5.19%. Moreover, we found that the benefits from having flexibility of price changes (dynamic discounts/premiums pricing model) and demand learning can complement each other to achieve as much as 8.15% revenue increase.

4.6 Conclusions

While revenue management (RM) has been intensively studied in the travel industries, Sport and Entertainment (S&E) also has a great potential to be improved by this idea but still has not obtained much attention in the literature. In addition, there maybe models or insights from the S&E industry that may apply to other RM

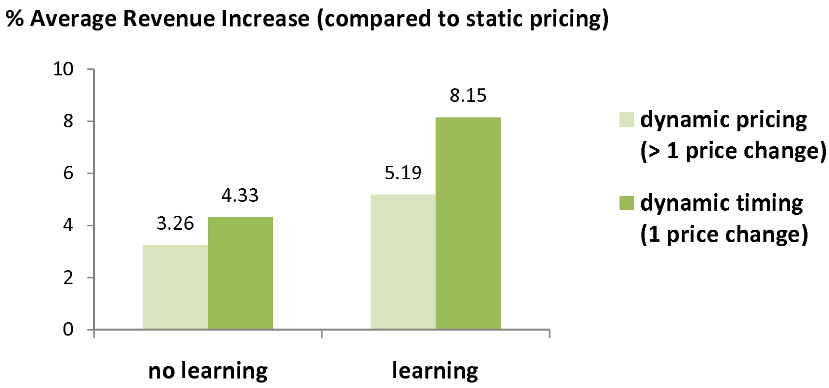

Figure 40: The average percentage revenue increase (from static pricing policy) under different scenarios

contexts. In this chapter, we developed pricing and timing models for stochastic S&E ticket demand in the existence of demand learning with Bayesian updates to reduce uncertainty and improve forecast. We also allow flexibility of ticket demand being affected by time since in the S&E industry, we observed significantly higher sales when it is closer to the end of the selling horizon.

We considered two types of models. The first model allows ticket price to be adjusted every period, where the adjustment is in terms of either the percentage increase (price premiums) or the percentage decrease (price discounts) with respect to the base price that has been determined at the beginning of the season. We found that the optimal ticket price discount/premium for the upcoming period is non-increasing with the amount of remaining inventory observed at the beginning of the period. To identify the benefit of demand learning, we compared the described model to a similar model without learning. We observed that demand learning is most beneficial when the initial demand estimation is inaccurate since the seller can correct his belief after observing the actual sales (with up to 8-11% improvement in revenue when the misestimation is high). In addition, our results showed: when incorporating demand learning, the underestimation of the base demand rate causes less revenue loss as compared to overestimation. This implies a risk-averse ticket seller who tends to underestimate the demand rate may obtain higher revenue than a risk-taker seller who overestimates it.

Instead of allowing price changes every period, our second pricing model limits the frequency of discounts/premiums to only once since customer satisfaction can be important for some sellers. After the seller decides to use a discount or premium, it will be applied in all remaining periods. So, the seller's decisions are the optimal timing to offer a premium/discount and how much. This model is suitable when changing price too often is not preferred, e.g., when frequent price change may reduce customer satisfaction or cause managerial complicatedness. We found that it is not always optimal to delay the discounts/premiums pricing decisions in order to learn more about demand. Our model's optimal decision is significantly related to the remaining inventory; i.e., the seller should offer discounts (premiums) when he observes the inventory is lower (higher) than a particular level. For moderate levels of inventory, the seller should delay his decision by continuing with the base price and reconsidering this decision in the next period.

We compared the performance of the second model (dynamic timing) with the first model (dynamic discounts/premiums pricing) and observed that when the standard deviation of the base demand rate distribution is small, the dynamic timing model can achieve a comparable revenue to that of the dynamic discounts/premiums pricing model. The models' performances were also found to be related to the timing effects on demand. When a larger amount of demand arrives closer to the end of the selling horizon, our results showed that it is less necessary to frequently change price. Moreover, we found that the benefits from having flexibility of price changes and demand learning can complement each other to achieve as much as 8.15% revenue increase on average, as compared to static pricing. We believe our model can encourage future revenue management model development for the Sport and Entertainment industry, and provide the basis for further research in dynamic pricing policies with demand learning since the context can be applied to other industries, selling perishable products with no inventory replenishment.

CHAPTER V

CONCLUDING REMARKS AND FUTURE WORKS

In this thesis, we presented innovative demand management techniques for service systems with limited resources. Our objective is to develop mathematical models to enhance the overall understanding of such systems and improve decisions, ultimately resulting in higher profit and better customer service. The first part of the dissertation (Chapter 2) explored different kinds of No-Kill operational policies in managing demand (coordination with other shelters, adoption and neutering) with the goal of providing insights on how each method helps reduce the killing rates (or reduce the job rejection rate of other service systems). For the coordination policy, if one shelter is full, animals can be diverted to the other shelters in the NK system. When each shelter has the same adoption rate, we proved that the coordinating NK system has a lower killing rate than the uncoordinated K system. For large NK systems, we found that adding more coordinated shelters helps reduce the overall killing probability but the marginal improvement is decreasing. We provided several insights on how to design different kinds of coordinating systems to achieve the highest improvement, including coordinating with neighbors or between particular pairs of shelters.

If the existing killing rates of all shelters are high, we found it is worthwhile to explore other alternatives such as neutering campaigns to reduce arrivals, and adoption campaigns to increase turnovers. When animals increase according to a type of growth function, and when a shelter promotes neutering, we showed the shelter's total cost is first concave and then convex with the number of animals neutered. For the adoption policy, if adoptions are concave and increasing with the number of the campaigns, we found the shelter's total cost is convex in the number of campaigns used and there is an optimal number of campaigns for the minimum cost. Our results showed that an adoption campaign can be more attractive for a low campaign budget, while a neutering policy is more effective for a higher budget when there are enough resources to neuter a sufficiently large number of animals. Although this work is motivated by the animal shelters' policies, the context can be applied to other industries with customers and servers, e.g., hospital operations, ambulances and internet servers. To the best of our knowledge, we are the first to quantitatively analyze operational policies of animal shelters or study them as a stochastic system.

For Chapter 2's future extensions, it would be interesting to consider other types of coordination for service systems. For example, instead of coordination based on the service rates of systems as considered in this chapter, it would be of interest to study other priority schemes such as coordination based on the cost of services and/or the cost of transferring customers. Empirical works on coordination in this area could be useful although it is difficult here as non-profit organizations rarely keep or provide detailed records. Systems can also change dramatically in a short time period (e.g., after promoting adoptions), one could explore how the rejection rates right after the policy implementation are compared to the rates after a certain time period, and how to effectively extend the reduced rejection rate effects.

In the second part of the thesis (Chapter 3), we considered a topic of demand management via "Scaling the house" decisions for the performance ticket industry, i.e., how to divide seats into zones for different prices to maximize revenue. We obtained transactional level data of over 300 shows from a major performance arts organization in the United States. We found that in addition to price, demand is significantly driven by the distance from the front and the distance from the center. We developed a two-dimensional "Scaling the House" model and identified the optimal closed form solutions of the number of seating rows (from the front) and the number of seats (from the center of rows) to be charged at a premium price that maximize the overall revenue. We developed an alternative one-dimensional zoning model, where the optimal seating zones are defined only by row cuts, for venues whose ticket demand does not significantly decrease with the distance from the center. We showed the optimal cutting point occurs at the seating row whose expected revenue when charging at a high price is equal to its expected revenue when charging at a low price.

We performed comparative statics of the two-price zoning decisions and showed that the optimal number of high-price seating rows decreases with customer price sensitivity. Thus, if customers are very sensitive to high prices, the venue managers should provide a smaller premium section and induce more revenue with more low-priced seats. Substitution effects can also impact the optimal zoning decision. We found that if customers are likely to switch from the front section (with high price) to the back section for a lower price, the it can be useful to move the low-price section further from the stage. For location sensitivity, when demand decreases significantly with distance from the front, it can be beneficial to move the low-price section closer to the stage. Our results also imply that when different types of performances are performed in the same venue, the management should consider having different seating charts for each show type. Further, we found that different venue's shape can imply different optimal zoning models and we provided insights on how to select appropriate seating sections, based on the venues' size and shape. Our work sheds light on "scaling the house" strategies that to date have largely been performed visually or based on ad-hoc basis.

Several directions are possible for future work related to Chapter 3. With the advancement of real-time sales information, some venues may consider dynamic zoning decisions, such as offering sub-sections during the selling season for some seating areas having low to-date sales, although this could be achieved through pricing promotions. Also, one could extend the model for dynamic decisions, e.g., when to change the seating chart, how and which seating section to re-scale.

In the third part of the dissertation (Chapter 4), we developed pricing and timing models for stochastic Sports and Entertainment (S&E) ticket demand in the existence of demand learning. Ultimately the goal is to improve price decisions and increase revenue. Unlike the existing works in demand learning area, our model assumes ticket demand is also affected by time because in the S&E industry, we observed a significantly higher sales when it is closer to the end of the selling horizon. Since the show's popularity is usually uncertain to the seller, we propose a method to learn the overall popularity via Bayesian updates as early sales are revealed.

We considered two situations: (1) the seller can adjust ticket prices every period, or (2) the seller dynamically determines the optimal timing and amount of a single price change in the selling horizon. Our results showed that demand learning is most beneficial when the initial estimates are incorrect (with up to 8-11% improvement in revenue when the misestimation is high). Our model's optimal decision is significantly related to the remaining inventory; i.e., in experiments, the seller should offer discounts (premiums) when he observed the inventory is lower (higher) than a particular level. For moderate levels of inventory, the seller should delay his decision by continuing with the base price and reconsidering this decision in the next period. We found it is less necessary for the seller to vary price every period if demand variation is low and/or a large amount of demand arrives close to the show dates. Moreover, our results showed that changing price every period is less necessary when demand learning is not allowed, while the benefit of frequent price changing increases if the seller updates his belief via the proposed demand learning procedure. Specifically, we found that the benefits from having flexibility of price changes and demand learning can complement each other to achieve as much as 8.15% revenue increase on average, as compared to static pricing.

There are several possible extensions of Chapter 4. First, our model assumes demand uncertainty comes from the base demand rate, while the price and the timing effects are unchanging. One could extend the model to allow imperfect information on price and/or timing sensitivities to explore the form of the resulting posterior distribution. We expect the distribution will be more complex in those cases, but it could be interesting to examine if the benefit can offset the higher complexity. A second promising extension is a pricing model with demand learning for substitutable products, e.g., substitutable seating sections or shows. If the ticket seller would like to allow different discounts/premiums for different types of seats or different days of the same performance, it can be useful to study how he can employ the observed sales to predict future demand for a variety of substitutable products (although there may not exist a closed form posterior distribution for each of them). Since dynamic pricing is fairly new for the S&E business, an empirical work exploring the short-term and long-term effects of this revenue management idea could be useful. Considering strategic customers for the S&E industry could also be a possible extension. Obviously, there is still room for revenue management improvement for the Sports and Entertainment ticket industry and we hope our study will encourage future research in this area.

Overall, we considered innovation policies to manage demand towards increasing profit or improving customer service in service systems, which are key components in our societies. We explored a variety of decision levels from planning to real-time, and we considered policies including partial coordination, segmentation, and decision making incorporating demand learning. In all cases, we focused on problems specific to a particular type of organizations but we also contributed new results across organizations and sectors.

APPENDIX A

APPENDIX FOR CHAPTER 2

Proof for Theorem 1

Proof. We refer to Figure 2 for the following proof. For $m = 2$, let $p(x, y)$ be the long-run probability of being in state (x, y). From the balance equation of state (0,0),

$$\begin{aligned} 2\lambda p(0,0) &= \mu[p(0,1) + p(1,0)], \\ \Rightarrow p(0,1) + p(1,0) &= \frac{2\lambda}{\mu} p(0,0). \end{aligned}$$

For states (0,1) and (1,0), the balance equations are:

$$\begin{aligned} (2\lambda + \mu)p(0,1) &= \lambda p(0,0) + \mu p(1,1) + 2\mu p(0,2) \\ (2\lambda + \mu)p(1,0) &= \lambda p(0,0) + \mu p(1,1) + 2\mu p(2,0). \end{aligned}$$

From the three equations above, we have

$$\begin{aligned} p(0,2) + p(1,1) + p(2,0) &= \frac{\lambda}{\mu}[p(0,1) + p(1,0)] \\ &= \frac{[2\lambda]^2}{2!\mu^2} p(0,0). \end{aligned}$$

Considering states (0,2),(1,1) and (2,0), the balance equations are given by

$$\begin{aligned} (2\lambda + 2\mu)p(0,2) &= \lambda p(0,1) + \mu p(1,2) + 3\mu p(0,3), \\ (2\lambda + 2\mu)p(1,1) &= \lambda[p(0,1) + p(1,0)] + 2\mu p(1,2) + 2\mu p(2,1), \\ (2\lambda + 2\mu)p(2,0) &= \lambda p(1,0) + \mu p(2,1) + 3\mu p(3,0). \end{aligned}$$

Summing up three equations above, we have

$$\begin{aligned} p(0,3) + p(1,2) + p(2,1) + p(3,0) &= \frac{2\lambda}{3\mu}(p(0,2) + p(1,1) + p(2,0)) \\ &= \frac{[2\lambda]^3}{3!\mu^3} p(0,0). \end{aligned}$$

Likewise, we have

$$\begin{aligned}
p(1,3)+p(2,2)+p(3,1) &= \frac{[2\lambda]^4}{4!\mu^4}p(0,0)\\
p(2,3)+p(3,2) &= \frac{[2\lambda]^5}{5!\mu^5}p(0,0)\\
p(3,3) &= \frac{[2\lambda]^6}{6!\mu^6}p(0,0).
\end{aligned}$$

Therefore, it is easy to see that the probability of having n animals in the entire NK system is $p(n) = \sum_{x,y:x+y=n} p(x,y)$, where $p(n) = \frac{[\frac{2\lambda}{\mu}]^n}{n!}p(0,0)$. This is the same as the probability of having n animals in an $M/M/2C/2C$ queue with the total arrival rate, 2λ, and the service rate, μ. Thus, we have

$$P_{2C}^{NK} = \frac{\frac{(\frac{2\lambda}{\mu})^{2C}}{2C!}}{\sum_{i=0}^{2C}\frac{(\frac{2\lambda}{\mu})^i}{i!}}.$$

Let us next consider the situation when there are m shelters in the system.

For $m \geq 2$, the proof is similar to the argument above where the dimension of state spaces ≥ 2. Thus, for an NK system with m shelters, we have

$$P_{mC}^{NK} = \frac{\frac{(\frac{m\lambda}{\mu})^{mC}}{(mC)!}}{\sum_{i=0}^{mC}\frac{(\frac{m\lambda}{\mu})^i}{i!}},$$

which is equal to the killing (rejecting) probability of an $M/M/mC/mC$ system. With the total equal arrival rate, $m\lambda$, the overall killing rates of the NK with m shelters and the $M/M/mC/mC$ system are also the same. ■

Proof for Theorem 1 (in a general case where shelters have different arrival rates and capacities)

Proof. Consider an NK system with two shelters (presented by Figure 41), where the arrival rates of the first and the second shelters are λ_1 and λ_2, respectively, and their capacities are C_1 and C_2, correspondingly.

From the balance equations of state (0,0),we have:

$$\begin{aligned}
2(\lambda_1+\lambda_2)p(0,0) &= \mu[p(0,1)+p(1,0)],\\
\Rightarrow p(0,1)+p(1,0) &= \frac{2(\lambda_1+\lambda_2)}{\mu}p(0,0). \qquad (32)
\end{aligned}$$

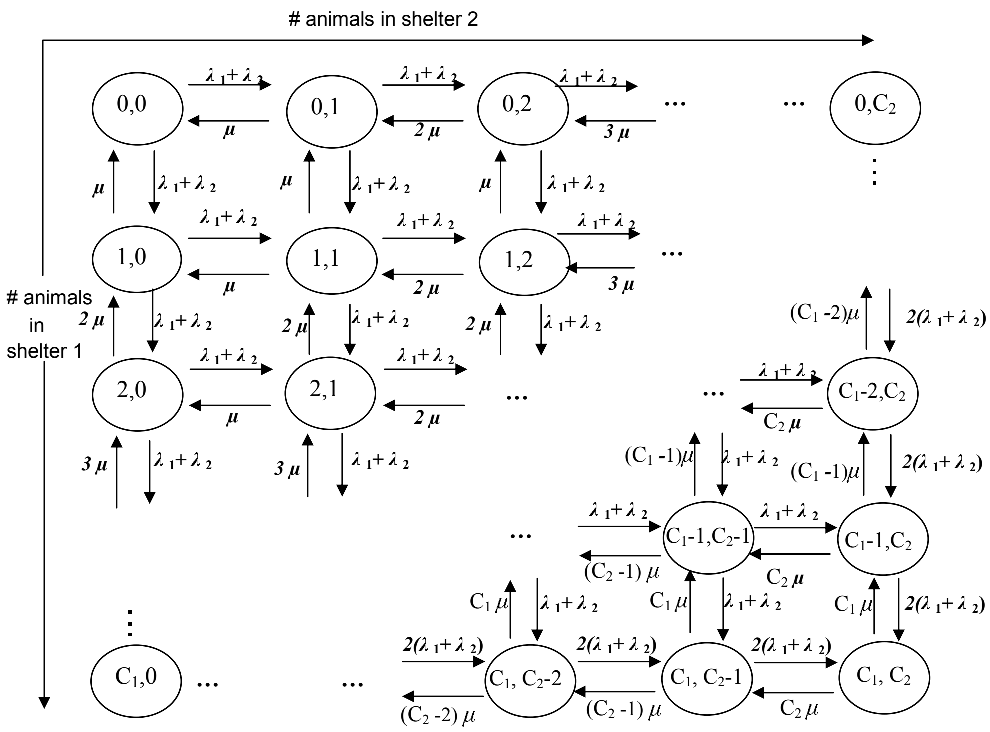

Figure 41: An NK system of two shelters with different arrival rates and capacities

Considering the balance equations of states (0,1) and (1,0), respectively, we have:

$$(2(\lambda_1+\lambda_2)+\mu)p(0,1) = \lambda_2 p(0,0)+\mu p(1,1)+2\mu p(0,2) \tag{33}$$

$$(2(\lambda_1+\lambda_2)+\mu)p(1,0) = \lambda_1 p(0,0)+\mu p(1,1)+2\mu p(2,0). \tag{34}$$

From the equations (32), (33) and (34), it follows that:

$$\begin{aligned} p(0,2)+p(1,1)+p(2,0) &= \frac{(\lambda_1+\lambda_2)}{2\mu}[p(0,1)+p(1,0)] \\ &= \frac{[(\lambda_1+\lambda_2)]^2}{2!\mu^2}p(0,0). \end{aligned}$$

Considering the balance equations of states (0,2),(1,1) and (2,0), respectively, we have:

$$\begin{aligned} ((\lambda_1+\lambda_2)+2\mu)p(0,2) &= \lambda_2 p(0,1)+\mu p(1,2)+3\mu p(0,3), \\ ((\lambda_1+\lambda_2)+2\mu)p(1,1) &= \lambda_1 p(0,1)+\lambda_2 p(1,0)+2\mu p(1,2)+2\mu p(2,1), \\ ((\lambda_1+\lambda_2)+2\mu)p(2,0) &= \lambda_1 p(1,0)+\mu p(2,1)+3\mu p(3,0). \end{aligned}$$

Summing up three equations above,

$$\begin{aligned} p(0,3)+p(1,2)+p(2,1)+p(3,0) &= \frac{(\lambda_1+\lambda_2)}{3\mu}(p(0,2)+p(1,1)+p(2,0)) \\ &= \frac{[(\lambda_1+\lambda_2)]^3}{3!\mu^3}p(0,0). \end{aligned}$$

Likewise, we have:

$$\begin{aligned} p(C_1-2,C_2)+p(C_2-1,C_2-1)+p(C_1,C_2-2) &= \frac{[(\lambda_1+\lambda_2)]^{C_1+C_2-2}}{(C_1+C_2-2)!\mu^{C_1+C_2-2}}p(0,0) \\ p(C_1-1,C_2)+p(C_1,C_2-1) &= \frac{[(\lambda_1+\lambda_2)]^{C_1+C_2-1}}{(C_1+C_2-1)!\mu^{C_1+C_2-1}}p(0,0) \\ p(C_1,C_2) &= \frac{[(\lambda_1+\lambda_2)]^{C_1+C_2}}{(C_1+C_2)!\mu^{C_1+C_2}}p(0,0). \end{aligned}$$

Thus, the probability of having n animals in an NK system with two shelters, where the first (second) shelter's arrival rate and capacity are λ_1 and C_1 (λ_2 and C_2), is $p(n) = \sum_{x,y:x+y=n} p(x,y){=}p(n) = \frac{[\frac{(\lambda_1+\lambda_2)}{\mu}]^n}{n!}p(0,0)$. This is the same as the probability of having n animals in an $M/M/(C_1+C_2)/(C_1+C_2)$ queue with the total arrival rate, $\lambda_1+\lambda_2$, and the service rate, μ. Consequently, it follows that

$$P^{NK}_{C_1+C_2} = \frac{\frac{(\frac{(\lambda_1+\lambda_2)}{\mu})^{C_1+C_2}}{(C_1+C_2)!}}{\sum_{i=0}^{C_1+C_2}\frac{(\frac{(\lambda_1+\lambda_2)}{\mu})^i}{i!}}.$$

We note that the proof for $m \geq 2$ (where each shelter j's arrival rate and capacity equal λ_j and C_j, $j = 1, ..., m$) is similar to the argument above. Thus,

$$P^{NK}_{(\sum_{j=1}^m C_j)} = \frac{\frac{(\frac{\sum_{j=1}^m \lambda_j}{\mu})^{\sum_{j=1}^m C_j}}{(\sum_{j=1}^m C_j)!}}{\sum\limits_{i=0}^{\sum_{j=0}^m C_j}\frac{(\frac{\sum_{j=1}^m \lambda_j}{\mu})^i}{i!}},$$

which is equal to the killing (rejecting) probability of an $M/M/(\sum_{j=1}^m C_j)/(\sum_{j=1}^m C_j)$ system. With the total equal arrival rate, $\sum_{j=1}^m \lambda_j$, the overall killing rates of the NK with m shelters and the $M/M/(\sum_{j=1}^m C_j)/(\sum_{j=1}^m C_j)$ system also equal. ■

Proof for Corollary 1

Proof. Let $P^{M/M/mC/mC}$ be the killing probability of the $M/M/mC/mC$ system having total arrival rate $m\lambda$. From Theorem 1, we have $P^{NK}_{mC} = P^{M/M/mC/mC}$. Since

$$P^{M/M/mC/mC} = \frac{1}{\int_0^\infty e^{-x}(1+\frac{x}{\frac{m\lambda}{\mu}})^{mC}dx},$$

and $(1 + \frac{x}{\frac{m\lambda}{\mu}})^{mC}$ is strictly increasing in m, we have that the rejecting probability of $M/M/mC/mC$, $P^{M/M/mC/mC}$, is strictly decreasing in m (see [68], and [131] for the proofs). Thus, $P^{NK}_{mC} = P^{M/M/mC/mC} < P^{M/M/C/C} = P^K_C$. Since both NK system and K systems with m shelters have total arrival rates $m\lambda$, it follows that the killing rate of the NK system is less than the rate of the K system. ■

Proof for Property 1

Proof. We have

$$\frac{\partial P^K_{C,2-type}}{\partial \lambda_1} = \frac{\partial P^K_{C,2-type}}{\partial r}\frac{\partial r}{\partial \lambda_1} = \frac{\partial P^K_{C,2-type}}{\partial r}\frac{1}{\mu_1}$$

$$\frac{\partial P^K_{C,2-type}}{\partial \lambda_2} = \frac{\partial P^K_{C,2-type}}{\partial r}\frac{\partial r}{\partial \lambda_2} = \frac{\partial P^K_{C,2-type}}{\partial r}\frac{1}{k\mu_1} > \frac{\partial P^K_{C,2-type}}{\partial r}\frac{1}{\mu_1} = \frac{\partial P^K_{C,2-type}}{\partial \lambda_1},$$

since $0 < k < 1$ and $\frac{\partial P^K_{C,2-type}}{\partial r} \geq 0$ (see [64] for the proof of these inequalities).

Proof for Property 2

Proof. We have

$$\frac{\partial P^K_{C,2-type}}{\partial k} = \frac{\partial P^K_{C,2-type}}{\partial r}\frac{\partial r}{\partial k} = \frac{\partial P^K_{C,2-type}}{\partial r}(-\frac{\lambda_2}{k^2\mu_1}) < 0$$

($\frac{\partial P^K_{C,2-type}}{\partial r} \geq 0$. ■

Proof for Theorem 2

Proof. Consider an NK system consisting of m identical shelters with two types of animals (i.e., the coordinated m $M/G/C/C$ systems). If there are n total animals in the entire NK system, define the state $T = (t_1, t_2, ..., t_n), t_1 \leq t_2 \leq ... \leq t_n$, when the most recent animal has arrived t_1 time units ago, the next most recent animal has arrived t_2 time units ago, and so on. We shall use the reverse process to find the limiting probability density $p(t_1, t_2, ..., t_n), 1 \leq n \leq mC, t_1 \leq t_2 \leq \cdots \leq t_n$. Let $p(\emptyset)$ be the probability that there are no animals in the NK system. As defined, the NK system with m shelters has total arrival rate $m(\lambda_1+\lambda_2)$. Since the mean adoption time of each shelter is identical, the overall inter-adoption time has general distribution G. Suppose that G has density g and failure rate function $a(t)$, where $a(t) = \frac{g(t)}{\overline{G}(t)}$ and $\overline{G}(t) = \int_t^\infty g(t)dt$. In other words, $a(t)$ is the instantaneous probability intensity that a t-unit-old animals in any shelters is adopted.

The goal is to prove the conjecture that the reverse process is also a coordination of m $M/G/C/C$ systems where arrivals follow a Poisson process with rate $m(\lambda_1 + \lambda_2)$. The state at anytime represents the ordered residual adoption times of animals currently in the NK system.

We will prove the above conjecture and also obtain the limiting distribution. For any state $T = (t_1, ..., t_i, ..., t_n)$, let $e_i(T) = (t_1, ..., t_{i-1}, t_{i+1}, ..., t_n)$. The probability density in the forward process of moving from state $T = (t_1, t_2, ..., t_n), 1 \leq n \leq mC$ to state $e_i(T)$ is equal to $a(t_i)$ since the animal that has spent t_i time units old must be suddenly adopted. If we consider the reversed process, the joint probability density of moving from state $e_i(T)$ to T is equal to $m(\lambda_1 + \lambda_2)g(t_i)$, since we need to have an animal with inter-adoption time t_i to suddenly enter the NK system. Thus, for a process to be reversible, we need to have that

$$\begin{aligned} p(T)a(t_i) &= p(e_i(T))m(\lambda_1 + \lambda_2)g(t_i) \\ p(T) &= p(e_i(T))m(\lambda_1 + \lambda_2)\overline{G}(t_i), \end{aligned}$$

which is true since $a(t) = \frac{g(t)}{\overline{G}(t_i)}$.

Let i=1. We have,

$$\begin{aligned} p(T) &= m(\lambda_1 + \lambda_2)\overline{G}(t_1)p(e_1(T)) \\ &= m(\lambda_1 + \lambda_2)\overline{G}(t_1)m(\lambda_1 + \lambda_2)\overline{G}(t_2)p(e_1(e_1(T))) \\ &\vdots \\ &= \prod_{i=1}^{n} m(\lambda_1 + \lambda_2)\overline{G}(t_i)p(\emptyset). \end{aligned} \tag{35}$$

To find $p(n\ animals\ in\ the\ NK)$, integrate over all T:

$$\begin{aligned} p(n\ animals\ in\ the\ NK) &= p(\emptyset)(m(\lambda_1+\lambda_2))^n \int\int\cdots\int_{t_1\le t_2\le\ldots\le t_n} \prod_{i=1}^{n} \overline{G}(t_i)dt_1dt_2\cdots dt_n \\ &= p(\emptyset)\frac{(m(\lambda_1+\lambda_2))^n}{n!}\int\int\cdots\int_{t_1,t_2,\ldots,t_n}\prod_{i=1}^{n}\overline{G}(t_i)dt_1dt_2\cdots dt_n \\ &= p(\emptyset)\frac{(m(\lambda_1+\lambda_2)E[S])^n}{n!}, n=1,2,...mC, \end{aligned}$$

where $E[S]=\int\overline{G}(t)dt$, which is the mean inter-adoption time. Thus, we have

$$p(n\ animals\ in\ the\ NK) \;=\; \frac{\frac{(m(\lambda_1+\lambda_2)E[S])^n}{n!}}{\sum_{i=0}^{mC}\frac{(m(\lambda_1+\lambda_2)E[S])^i}{i!}}, n=0,1,...,mC. \tag{36}$$

From (1), we have,

$$p(T) \;=\; \frac{((m(\lambda_1+\lambda_2))^n\prod_{i=1}^{n}\overline{G}(t_i)}{\sum_{i=0}^{mC}\frac{(m(\lambda_1+\lambda_2)E[S])^i}{i!}}. \tag{37}$$

The conditional distribution of the ordered ages given that there are n animals in the NK system is

$$p(T\mid n\ animals\ in\ the\ NK)=\frac{p(T)}{p(n\ animals\ in\ the\ NK)}=n!\prod_{i=1}^{n}\frac{\bar{G}(t_i)}{E[S]}.$$

$\frac{\bar{G}(t_i)}{E[S]}$ is the density of the equilibrium distribution of G, G_e. To have the above conjecture valid, we can see from (2) that the limiting distribution of the number in the system only depends the mean adoption time, and if there are n animals in the NK system, their ages are independent and identically distributed as the equilibrium distribution G_e of G.

In addition, we need to consider the forward process moving from T to $(0,t_1,...,t_n)$ when $n<k$. The probability intensity in that case is $m(\lambda_1+\lambda_2)$. On the other hand, in the reverse process, moving from $(0,t_1,...,t_n)$ to T occurs with probability 1. Thus, for a process to be reversible, we need that

$$p(T)m(\lambda_1+\lambda_2)=p(0,t_1,...,t_n),$$

which is true since it follows from (3) and $\bar{G}(0) = 1$. Therefore, we have that the reverse process of the NK system consisting of m shelters (coordinated m $M/G/C/C$ systems) also has the inter-adoption distribution G, and arrival process is also a Poisson process with rate $m(\lambda_1 + \lambda_2)$. Moreover, the state at any time represents the ordered residual inter-adoption times of animals in the NK system. The limiting distribution of the number of animals depends on G only through its mean, and their (unordered) ages are i.i.d. due to the equilibrium distribution G_e of G. Thus, the rejecting probability of the NK system with two type of animals (the coordinated m $M/G/C/C$) is equal to that of the $M/G/mC/mC$ system (which is equal to the rejecting probability of the $M/M/mC/mC$). Thus, the NK system with two types gives the same killing probability as the NK system with one type of animal. ■

Proof for Lemma 1

Proof.

i) From Little's Law ([93]), $L = \lambda_{eff} W$, where L, λ_{eff} and W are the long run number of customers, the effective arrival rates, and the average time each customer spending in the system, respectively. Since we assume there is no waiting space in the shelter, $W = \frac{1}{\nu}$. Thus, in the NK system with m shelters, we have

$$L^{NK}_{mC} = m(\lambda_1 + \lambda_2)(1 - P^{NK}_{mC})\frac{1}{\nu}.$$

In the K system with m shelters,

$$L^{K}_{mC} = m(\lambda_1 + \lambda_2)(1 - P^{K}_{C})\frac{1}{\nu}.$$

From Corollary 1, $P^{NK}_{mC} < P^{K}_{C}$. Therefore, $L^{NK}_{mC} > L^{K}_{mC}$. ■

ii) From Theorem 1, the killing rates of the NK and $M/G/mC/mC$ are equal. Also, from the proof in Smith and Whitt (1980), the rejecting probability of the

$M/G/mC/mC$ system is decreasing in m. Therefore, it follows that the killing probability of the system with m shelters under the NK policy is decreasing in m.

Proof for Property 3

Proof.

i) Since $\lim_{r\to 0} P_C^K = \lim_{r\to 0} \frac{\frac{r^C}{C!}}{\sum_{i=0}^{C} \frac{r^i}{i!}} = 0$ and $\lim_{r\to 0} P_{mC}^{NK} = \lim_{r\to 0} \frac{\frac{(mr)^{(mC)}}{(mC)!}}{\sum_{i=0}^{mC} \frac{(mr)^i}{i!}} = 0$.

ii) Since $\lim_{r\to \infty} P_C^K = \lim_{r\to \infty} \frac{\frac{r^C}{C!}}{\sum_{i=0}^{C} \frac{r^i}{i!}} = 1$ and $\lim_{r\to \infty} P_{mC}^{NK} = \lim_{r\to \infty} \frac{\frac{(mr)^{(mC)}}{(mC)!}}{\sum_{i=0}^{mC} \frac{(mr)^i}{i!}} = 1$.

iii) Follows from Lemma 1 (ii).

vi) Follows from $\frac{\partial P_C^K}{\partial r} = P_C^K(\frac{C}{r} - 1 + P_C^K)$ and $\frac{\partial P_{mC}^{NK}}{\partial r} = P_{mC}^{NK}(\frac{C}{r} - 1 + P_{mC}^{NK})$. For detailed derivations, see Whitt (2002) ■

Proof for Observation 13

Proof. Consider the following example when m=2. The $M/G/2C/2C$ system consisting of two shelters as described has the total arrival rate, $2(\lambda_1 + \lambda_2)$ and the mean inter-adoption time, $\frac{\frac{\lambda_1}{\mu_{1,1}}+\frac{\lambda_2}{\mu_{2,1}}+\frac{\lambda_2}{\mu_{1,2}}+\frac{\lambda_2}{\mu_{2,2}}}{2(\lambda_1+\lambda_2)}$. Thus,

$$P^{M/G/2C/2C} = \frac{\frac{d^{2C}}{(2C)!}}{\sum_{i=0}^{2C} \frac{d^i}{i!}},$$

where $d = \frac{2(\lambda_1+\lambda 2)}{(\frac{2(\lambda_1+\lambda 2)}{\frac{\lambda_1}{\mu_{1,1}}+\frac{\lambda_2}{\mu_{2,1}}+\frac{\lambda_1}{\mu_{1,2}}+\frac{\lambda_2}{\mu_{2,2}}})} = \frac{\lambda_1}{\mu_{1,1}} + \frac{\lambda_2}{\mu_{2,1}} + \frac{\lambda_1}{\mu_{1,2}} + \frac{\lambda_2}{\mu_{2,2}}$.

Let $\mu_{1,1} = \phi$, $\mu_{2,1} = k\phi$, $\mu_{1,2} = \frac{1}{\phi}$, and $\mu_{2,2} = \frac{k}{\phi}$, where $0 < k < 1$; then, as $\phi \to 0$, we have $d \to \infty$ and $P^{M/G/2C/2C} \to 1$.

On the other hand, consider an NK system consisting of the two shelters as described. As $\phi \to 0$, we have $\mu_{1,2} \to \infty$, and $\mu_{2,2} \to \infty$. The rejected animals from shelter 1 can always be diverted to shelter 2 and get adopted there and we have $P_{2C}^{NK} \to 0$ for this case. Thus, if the mean adoption rate of each shelter in the system is different, the total killing rate of the NK system can be different than the rate of the $M/G/mC/mC$ system. ■

Proof for Lemma 2

Proof. To prove this lemma, we show that if $\frac{\partial^2 \mu_1(n_a)}{\partial n_a^2} \leq 0$, then $\frac{\partial^2 TC_a^{NK}(n_a)}{\partial n_a^2} \geq 0$. Considering the total cost,

$$\frac{\partial TC_a^{NK}(n_a)}{\partial \mu_1(n_a)} = v_k(\lambda_1+\lambda_2)\frac{\partial P_C(\mu_1(n_a))}{\partial \mu_1(n_a)}.$$

From the chain rule ([144]), the second derivative of a shelter's cost with respect to the number of campaigns can be written as

$$\frac{\partial^2 TC_a^{NK}(n_a)}{\partial n_a^2} = \frac{\partial^2 TC_a^{NK}(n_a)}{\partial \mu_1(n_a)^2}(\frac{\partial \mu_1(n_a)}{\partial n_a})^2 + \frac{\partial TC_a^{NK}(n_a)}{\partial \mu_1(n_a)}\frac{\partial^2 \mu_1(n_a)}{\partial n_a^2}.$$

From [64], $\frac{\partial P_C(\mu_1(n_a))}{\partial \mu_1(n_a)} < 0$ and $\frac{\partial^2 P_C(\mu_1(n_a))}{\partial \mu_1(n_a)^2} > 0$, thus we have $\frac{\partial TC_a^{NK}(n_a)}{\partial \mu_1(n_a)} < 0$ and $\frac{\partial^2 TC_a^{NK}(n_a))}{\partial \mu_1(n_a)^2} > 0$. Since $(\frac{\partial \mu_1(n_a)}{\partial n_a})^2 \geq 0$, we have $\frac{\partial^2 TC_a^{NK}(n_a)}{\partial n_a^2} \geq 0$ if $\frac{\partial^2 \mu_1(n_a)}{\partial n_a^2} \leq 0$. ■

Proof for Theorem 3

Proof. The convexity shown in Lemma 2 allows us to determine the optimal value of n_a by setting $\frac{\partial TC_a^{NK}(n_a)}{\partial n_a}$ equal to 0. From equation (1),

$$\begin{aligned}\frac{\partial TC_a^{NK}(n_a)}{\partial n_a} &= v_k(\lambda_1+\lambda_2)\frac{\partial P_C(\mu_1(n_a))}{\partial \mu_1(n_a)}\frac{\partial \mu_1(n_a)}{\partial n_a} + v_a \\ &= -v_k(\lambda_1+\lambda_2)P_C(\mu_1(n_a))(\frac{C\mu_1(n_a)}{\tilde{\lambda}} - 1 + P_C(\mu_1(n_a)))\frac{\tilde{\lambda}}{\mu_1(n_a)^2}\frac{\partial \mu_1(n_a)}{\partial n_a} + v_a\end{aligned}$$

since $\frac{\partial P_C(\mu_1(n_a))}{\partial \mu_1(n_a)} = -P_C(\mu_1(n_a))(\frac{C\mu_1(n_a)}{\tilde{\lambda}} - 1 + P_C(\mu_1(n_a)))\frac{\tilde{\lambda}}{\mu_1(n_a)^2} < 0$ (see [64]). In case i) when $v_k(\lambda_1+\lambda_2)P_C(\mu_1(0))(\frac{C\mu_1(0)}{\tilde{\lambda}} - 1 + P_C(\mu_1(0)))\frac{\tilde{\lambda}}{\mu_1(0)^2}\frac{\partial \mu_1(0)}{\partial n_a} < v_a$, we have that $TC_a^{NK}(n_a)$ is increasing in $[0, \overline{n}_a]$ and $n_a^* = 0$. In case ii) when $v_k(\lambda_1 + \lambda_2)P_C(\mu_1(\overline{n}_a))(\frac{C\mu_1(\overline{n}_a)}{\tilde{\lambda}} - 1 + P_C(\mu_1(\overline{n}_a)))\frac{\tilde{\lambda}}{\mu_1(\overline{n}_a)^2}\frac{\partial \mu_1(\overline{n}_a)}{\partial n_a} > v_a$, $TC_a^{NK}(n_a)$ is decreasing in $[0, \overline{n}_a]$ and $n_a^* = \overline{n}_a$. Otherwise, n_a^* is the solution of (2). ■

Proof for Lemma 3

Proof. To prove this lemma, we show that if $\frac{\partial^2 \lambda_1(n_s)}{\partial n_s^2} \leq 0$, then $\frac{\partial^2 TC_s^{NK}(n_s)}{\partial n_s^2} \geq 0$.

$$\text{Let } A_C = TC_s^{NK}(n_s) - (f_s + v_s n_s) = v_k(h\lambda(n_s) + \lambda(n_s))P_C(\lambda(n_s)).$$

From [32], we know that P_C has the recurrence-relation, $P_C(\lambda(n_s)) = \frac{rP_{C-1}(\lambda(n_s))}{C+rP_{C-1}(\lambda(n_s))}$.

Thus,

$$A_C = \frac{Qv_k(h+1)\lambda(n_s)A_{C-1}}{C + QA_{C-1}} \quad ,C \geq 1 \tag{38}$$

where $Q = \frac{\frac{h}{\mu_1}+\frac{1}{k\mu_1}}{v_k(h+1)}$. It is easy to see that $A_0 = v_k(h+1)\lambda(n_s)$. It follows that:

$$A_0^{'} = \frac{\partial A_0}{\partial n_s} = v_k(h+1)\frac{\partial \lambda(n_s)}{\partial n_s},$$

$$A_0^{''} = \frac{\partial^2 A_0}{\partial n_s^2} = v_k(h+1)\frac{\partial^2 \lambda(n_s)}{\partial n_s^2},$$

for C=0. Thus, we have that $A_0^{'} \leq 0$ since $\lambda(n_s)$ is non-increasing in n_s ($\lambda^{'}(n_s) \leq 0$).

Next, we prove by induction that the above equation, $A_C^{'} \leq 0$, holds for all $C \geq 0$. As the induction hypothesis assumes that $A_{C-1}^{'} \leq 0$.

From equation (38),

$$A_C(C + QA_{C-1}) = Qv_k(h+1)\lambda(n_s)A_{C-1},$$

and taking the derivative of both sides with respect to n_s yields

$$\begin{aligned}
(C + QA_{C-1})A_C^{'} &= Qv_k(h+1)\lambda^{'}(n_s)A_{C-1} + QA_{C-1}^{'}[v_k(h+1)\lambda(n_s) - A_C] \\
&= Qv_k(h+1)\lambda^{'}(n_s)A_{C-1} + Q(1 - P_C(\lambda(n_s)))v_k(h+1)\lambda(n_s)A_{C-1}^{'}.
\end{aligned}$$

Therefore, using the induction hypothesis, we have $A_C^{'} \leq 0$ for all $C \geq 0$.

Taking the derivative of both sides of equation (39) with respect to n_s again yields

$$\begin{aligned}
(C + QA_{C-1})A_C^{''} &= 2QA_{C-1}^{'}[v_k(h+1)\lambda^{'}(n_s) - A_C^{'}] + QA_{C-1}^{''}[v_k(h+1)\lambda(n_s) - A_C] \\
&\quad + Qv_k(h+1)A_{C-1}\lambda^{''}(n_s) \qquad (39) \\
&= 2QA_{C-1}^{'}[v_k(h+1)\lambda^{'}(n_s) - A_C^{'}] + QA_{C-1}^{''}(1 - P_C(\lambda(n_s)))v_k(h+1)\lambda(n_s) \\
&\quad + Qv_k(h+1)A_{C-1}\lambda^{''}(n_s).
\end{aligned}$$

Consider $A_C^{'} = \frac{\partial A_C}{\partial n_s} = \frac{\partial A_C}{\partial g}\frac{\partial g}{\partial r}\frac{\partial r}{\partial n_s}$, where $r = \frac{h\lambda(n_s)}{\mu_1} + \frac{\lambda(n_s)}{k\mu_1}$, $g = rP_C(\lambda(n_s))$ and $A_C = \frac{g}{Q}$. We have,

$$A_C^{'} = \lambda^{'}(n_s)v_k(h+1)\frac{\partial g}{\partial r} \geq \lambda^{'}(n_s)v_k(h+1),$$

since $\frac{\partial g}{\partial r} \leq 1$ and $\lambda^{'}(n_s) \leq 0$. Thus, $2QA_{C-1}^{'}[v_k(h+1)\lambda^{'}(n_s) - A_C^{'}] \geq 0$.
If $\lambda^{''}(n_s) \geq 0$, then $A_0^{''} \geq 0$. If we assume as the induction hypothesis, $A_{C-1}^{''} \geq 0$, we then have that $A_C^{''} \geq 0$ and $\frac{\partial^2 TC_s^{NK}(n_s)}{\partial n_s^2} = A_C^{''} \geq 0$.

Thus, if $\lambda(n_s)$ is convex in n_s, then the shelter cost ($TC_s^{NK}(n_s)$) is a convex function with respect to the number of neutered animals, n_s. ■

Proof for Property 4

Proof.

i) The proof follows directly from Lemma 3 and the property that $\lambda(n_s)$ is convex when $n_s \geq \frac{\log \frac{w_s}{z_s}}{l_s}$.

ii) Taking the second derivative of equation (3), we obtain

$$\begin{aligned}\frac{\partial^2 TC_s^{NK}(n_s)}{\partial n_s^2} &= v_k(h+1)\Big[(\frac{\partial \lambda(n_s)}{\partial n_s})^2(2\frac{\partial P_C(\lambda(n_s))}{\partial \lambda(n_s)} + \lambda(n_s)\frac{\partial^2 P_C(\lambda(n_s))}{\partial \lambda(n_s)^2}) \\ &\quad + \frac{\partial^2 \lambda(n_s)}{\partial n_s^2}[P_C(\lambda(n_s)) + \lambda(n_s)\frac{\partial P_C(\lambda(n_s))}{\partial \lambda(n_s)}]\Big]. \qquad (40)\end{aligned}$$

Setting (40) equal to zero, we have

$$\begin{aligned}0 &= (\frac{\partial \lambda(n_s)}{\partial n_s})^2\Big[2\frac{\partial P_C(\lambda(n_s))}{\partial \lambda(n_s)} + \lambda(n_s)\frac{\partial^2 P_C(\lambda(n_s))}{\partial \lambda(n_s)^2}\Big] \\ &\quad + \frac{\partial^2 \lambda(n_s)}{\partial n_s^2}\Big[P_C(\lambda(n_s)) + \lambda(n_s)\frac{\partial P_C(\lambda(n_s))}{\partial \lambda(n_s)}\Big]. \qquad (41)\end{aligned}$$

The first term of (40) is greater than or equal to zero since $(\frac{\partial \lambda(n_s)}{\partial n_s})^2 \geq 0$ and $(2\frac{\partial P_C(\lambda(n_s))}{\partial \lambda(n_s)} + \lambda(n_s)\frac{\partial^2 P_C(\lambda(n_s))}{\partial \lambda(n_s)^2} \geq 0$ (from [80], the rejection rate is increasing convex in the arrival rate). The sign of the second term depends on the sign of $\frac{\partial^2 \lambda(n_s)}{\partial n_s^2}$ since $P_C(\lambda(n_s)) + \lambda(n_s)\frac{\partial P_C(\lambda(n_s)}{\partial \lambda(n_s)} \geq 0$. Since $\frac{\partial^2 \lambda(n_s)}{\partial n_s^2}$ changes sign from negative to positive only once, setting (40) equals to zero, we have that there is at most one sign change for (40). So, there is at most one positive root for (40) as well (see Descartes' Rule in [57]). Thus, there exists at most one $n_s^* < \frac{\log \frac{w_s}{z_s}}{l_s}$

(otherwise, when $n_s \geq \frac{\log \frac{w_s}{z_s}}{l_s}$, from (i), $\frac{\partial^2 TC_s^{NK}(n_s)}{\partial n_s^2} \geq 0$) such that the shelter cost is concave when $n_s < n_s^*$ and convex when $n_s \geq n_s^*$. ■

APPENDIX B

APPENDIX FOR CHAPTER 3

Proof for Lemma 4

Proof. The revenue maximization problem given in (7) can be written as:

$$\begin{aligned}
\max_{0\leq\rho_1\leq r, 0\leq\eta_1\leq c} Rev^{2D}_{n=2} &= 2\Big(P_1\int_0^{\eta_1}\int_0^{\rho_1}(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_FD-\beta_CC)dDdC \\
&\quad + P_2[\int_{\eta_1}^{c}\int_0^{r}(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_FD-\beta_CC)dDdC \\
&\quad + \int_0^{\eta_1}\int_{\rho_1}^{r}(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_FD-\beta_CC)dDdC]\Big), \\
&= 2\Big(P_1\eta_1\rho_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\frac{\rho_1}{2}-\beta_C\frac{\eta_1}{2}) \\
&\quad + P_2[r(c-\eta_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r}{2}-\beta_C\frac{c+\eta_1}{2}) \\
&\quad + \eta_1(r-\rho_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r+\rho_1}{2}-\beta_C\frac{\eta_1}{2})]\Big), \qquad (42)
\end{aligned}$$

subject to:

$$\begin{aligned}
\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\frac{\rho_1}{2}-\beta_C\frac{\eta_1}{2} &\leq 1, \qquad (43) \\
r(c-\eta_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r}{2}-\beta_C\frac{c+\eta_1}{2}) & \\
+\eta_1(r-\rho_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r+\rho_1}{2}-\beta_C\frac{\eta_1}{2}) &\leq rc-\rho_1\eta_1. \qquad (44)
\end{aligned}$$

From Assumption 4, the total demand of seats when priced at P_2 are assumed to be less than the total capacity, so constraint (44) always holds with inequality. The Lagrangean is given by:

$$\begin{aligned}
L(\rho_1,\eta_1) &= 2\Big(P_1\eta_1\rho_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\frac{\rho_1}{2}-\beta_C\frac{\eta_1}{2}) \\
&\quad + P_2[r(c-\eta_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r}{2}-\beta_C\frac{c+\eta_1}{2}) \\
&\quad + \eta_1(r-\rho_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r+\rho_1}{2}-\beta_C\frac{\eta_1}{2})]\Big) \\
&\quad - \mu(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\frac{\rho_1}{2}-\beta_C\frac{\eta_1}{2}-1),
\end{aligned}$$

where the first-order condition with respect to ρ_1 is given by:

$$\begin{aligned}\frac{\partial L(\rho_1,\eta_1)}{\partial \rho_1} &= 2\eta_1[P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\rho_1-\beta_C\frac{\eta_1}{2}) \\ &\quad - P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\rho_1-\beta_C\frac{\eta_1}{2})]+\mu\frac{\beta_F}{2}.\end{aligned}$$

The second derivative of $L(\rho_1,\eta_1)$ with respect to ρ_1 is $\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \rho_1^2} = -2\eta_1\beta_F(P_1-P_2) < 0$. Similarly, the first derivative of $L(\rho_1,\eta_1)$ with respect to η_1 is

$$\begin{aligned}\frac{L(\rho_1,\eta_1)}{\partial \eta_1} &= 2\rho_1[P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\frac{\rho_1}{2}-\beta_C\eta_1) \\ &\quad - P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{\rho_1}{2}-\beta_C\eta_1)]+\mu\frac{\beta_C}{2}.\end{aligned}$$

The second derivative of $L(\rho_1,\eta_1)$ with respect to η_1 is $\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \eta_1^2} = -2\rho_1\beta_C(P_1-P_2) < 0$. Consider $\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \rho_1\partial\eta_1}$ and $\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \eta_1\partial\rho_1}$, we have that $\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \rho_1\partial\eta_1} = \frac{\partial^2 L(\rho_1,\eta_1)}{\partial \eta_1\partial\rho_1} = 2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\rho_1-\beta_C\eta_1)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\rho_1-\beta_C\eta_1)]$.

It is easy to check that $(\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \rho_1^2})(\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \eta_1^2}) - (\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \rho_1\partial\rho_2})(\frac{\partial^2 L(\rho_1,\eta_1)}{\partial \rho_2\partial\rho_1})$ is not always positive. Thus, determinants of the hessian of $L(\rho_1,\eta_1)$ do not alternate in sign and this revenue maximizing problem is not jointly concave in the decision variables ρ_1 and η_1. ■

Proof for Lemma 5

Proof. We start by first solving for the optimal ρ_1 as a function of η_1. Consider the revenue maximization problem given in (42), the constraint (43) can be written as $\rho_1 \geq \frac{\alpha-\beta_1P_1+\beta_{21}P_2-1}{\beta_F}$ and $\eta_1 \geq \frac{\alpha-\beta_1P_1+\beta_{21}P_2-1}{\beta_C}$ (from Assumption 5). Thus, the revenue maximization reduces to:

$$\begin{aligned}\max_{\max\{0,M_F\}\leq\rho_1\leq r} Rev^{2D}_{n=2} &= 2\Big(P_1\eta_1\rho_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\frac{\rho_1}{2}-\beta_C\frac{\eta_1}{2}) \\ &\quad + P_2\big[r(c-\eta_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r}{2}-\beta_C\frac{c+\eta_1}{2}) \\ &\quad + \eta_1(r-\rho_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r+\rho_1}{2}-\beta_C\frac{\eta_1}{2})\big]\Big),\end{aligned}$$

where $M_F = \frac{\alpha-\beta_1P_1+\beta_{21}P_2-1}{\beta_F}$.

Similar to the proof in Lemma 4, we have

$$\frac{\partial Rev_{n=2}^{2D}(\rho_1,\eta_1)}{\partial \rho_1} =$$
$$2\eta_1[P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\rho_1-\beta_C\tfrac{\eta_1}{2})-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\rho_1-\beta_C\tfrac{\eta_1}{2})],$$

and

$$\frac{\partial^2 Rev_{n=2}^{2D}(\rho_1,\eta_1)}{\partial \rho_1^2} = -2\eta_1\beta_F(P_1-P_2) < 0.$$

The concavity of $Rev_{n=2}^{2D}$ with respect to ρ_1 allows us to find the optimal value of ρ_1 by setting the first derivative of $Rev_{n=2}^{2D}$ is equal to zero. The optimal $\rho_1^*(\eta_1)$ is given by

$$\begin{aligned}\rho_1^*(\eta_1) &= \max\{\min\{0, M_F,\\ &\quad \frac{P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)-(P_1-P_2)\beta_C\frac{\eta_1}{2}}{\beta_F(P_1-P_2)}\}, r\}.\end{aligned}$$

The following cases provide conditions for each optimal solution of $\rho_1^*(\eta_1)$.

<u>**Case A.**</u>

i) If $\alpha-\beta_1P_1+\beta_{21}P_2 \le 1$, we have $M_F \le 0$. When $P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_C\frac{\eta_1}{2}) < P_2(\alpha-\beta_2P_2+\beta_{12})P_1-\beta_C\frac{\eta_1}{2})$, then $Rev_{n=2}^{2D}$ is decreasing in the interval $[0,r]$, so $\rho_1^*(\eta_1)=0$.

ii) If $\alpha-\beta_1P_1+\beta_{21}P_2 > 1$, we have $M_F > 0$. When $P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_FM_F-\beta_C\frac{\eta_1}{2}) < P_2(\alpha-\beta_2P_2+\beta_{12})P_1-\beta_FM_F-\beta_C\frac{\eta_1}{2})$, then $Rev_{n=2}^{2D}$ is decreasing in the interval $[M_F,r]$, so $\rho_1^*(\eta_1)=M_F$.

<u>**Case B.**</u> When $P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr-\beta_C\frac{\eta_1}{2}) > P_2(\alpha-\beta_2P_2+\beta_{12})P_1-\beta_Fr-\beta_C\frac{\eta_1}{2})$, then $Rev_{n=2}^{2D}$ is increasing in the interval $[\max\{0,M_F\},r]$, so $\rho_1^*(\eta_1)=r$.

<u>**Case C.**</u>

i) If $\alpha-\beta_1P_1+\beta_{21}P_2 \le 1$:

$$P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_C\tfrac{\eta_1}{2}) \geq P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_C\tfrac{\eta_1}{2}), \text{ and}$$
$$P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr-\beta_C\tfrac{\eta_1}{2}) \leq P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr-\beta_C\tfrac{\eta_1}{2}),$$

and we have that $Rev^{2D}_{n=2}$ is increasing and then decreasing in the interval $[0,r]$.

ii) If $\alpha-\beta_1P_1+\beta_{21}P_2>1$:

$$P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_FM_F-\beta_C\tfrac{\eta_1}{2}) \geq P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_FM_F-\beta_C\tfrac{\eta_1}{2});$$
$$\text{and } P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr-\beta_C\tfrac{\eta_1}{2}) \leq P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr-\beta_C\tfrac{\eta_1}{2}),$$

and we have that $Rev^{2D}_{n=2}$ is increasing and then decreasing in the interval $[M_F,r]$.

In both cases (i) and (ii), $\rho_1^*(\eta_1)=\frac{\alpha(P_1-P_2)-\beta_1P_1^2+\beta_2P_2^2-(\beta_{12}-\beta_{21})P_1P_2-(P_1-P_2)\beta_C\frac{\eta_1}{2}}{\beta_F(P_1-P_2)}$.

Thus from case A, B and C,

$$\rho_1^*(\eta_1)=\min\{\max\{0,M_F,\tfrac{\alpha(P_1-P_2)-\beta_1P_1^2+\beta_2P_2^2-(\beta_{12}-\beta_{21})P_1P_2-(P_1-P_2)\beta_C\frac{\eta_1}{2}}{\beta_F(P_1-P_2)}\},r\}.$$

■

Proof for Theorem 4

Proof. After solving for the optimal $\rho_1^*(\eta_1)$ as a function of η_1 (given in Lemma 5), we substitute $\rho_1^*(\eta_1)$ into the revenue function $Rev^{2D}_{n=2}$ to next solve for the optimal η_1^*. The following cases identify the shape of $Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)$ for each optimal $\rho_1^*(\eta_1)$ when $\alpha-\beta_1P_1+\beta_{21}P_2\leq 1$.

Case I) When $\rho_1^*(\eta_1)=0$, $Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)=2P_2cr(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r}{2}-\beta_C\frac{c}{2})$, and it is independent with η_1.

Case II) When $\rho_1^*(\eta_1)=r$, we have

$$\begin{aligned} Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1) &= 2\Big[P_1\eta_1r(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\frac{r}{2}-\beta_C\frac{\eta_1}{2}) \\ &\quad + P_2r(c-\eta_1)(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r}{2}-\beta_C\frac{c+\eta_1}{2})\Big]. \end{aligned}$$

Taking the second derivative of $Rev^{2D}_{n=2}(\rho_1^*(\eta_1), \eta_1)$ with respect to η_1: $\frac{\partial^2 Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1^2} = -2r\beta_C(P_1 - P_2) < 0$. Thus, $Rev^{2D}_{n=2}(\rho_1^*(\eta_1), \eta_1)$ is concave in η_1.

Case III) When $\rho_1^*(\eta_1) = \frac{\alpha(P_1-P_2)-\beta_1 P_1^2+\beta_2 P_2^2-(\beta_{12}-\beta_{21})P_1P_2-(P_1-P_2)\beta_C\frac{\eta_1}{2}}{\beta_F(P_1-P_2)}$, we substitute $\rho_1^*(\eta_1)$ into $Rev^{2D}_{n=2}(\rho_1^*(\eta_1), \eta_1)$ and then take the second derivative of $Rev^{2D}_{n=2}(\rho_1^*(\eta_1), \eta_1)$ with respect to η_1:

$$\begin{aligned}\frac{\partial^2 Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1^2} &= 2\frac{\beta_C}{\beta_F}[P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_C\frac{3}{4}\eta_1)\\ &\quad - P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_C\frac{3}{4}\eta_1)].\end{aligned}$$

We can see that $\frac{\partial^2 Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1^2}$ can be either positive or negative, depending on model parameters. The third derivative of $Rev^{2D}_{n=2}(\rho_1^*(\eta_1), \eta_1)$ is

$$\frac{\partial^3 Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1^3} = \frac{3(\beta_C)^2(P_1-P_2)}{2\beta_F} > 0.$$

So, $\frac{\partial Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1}$ has at most two roots. Moreover, since $\frac{\partial^3 Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1^3} > 0$, we have that $\frac{\partial^2 Rev^{2D}_{n=2}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1^2}$ changes from negative to positive. So, the revenue function $Rev^{2D}_{n=2}(\rho_1^*(\eta_1), \eta_1)$ is first concave and then convex in η_1.

From Lemma 5, when $\alpha - \beta_1P_1 + \beta_{21}P_2 \leq 1$, we have

$$\rho_1^*(\eta_1) = \Big(\min\{\frac{P_1(\alpha-\beta_1P_1+\beta_{21}P_2) - P_2(\alpha-\beta_2P_2+\beta_{12}P_1) - (P_1-P_2)\beta_C\frac{\eta_1}{2}}{\beta_F(P_1-P_2)}, r\}\Big)^+.$$

Next, we identify the optimal η_1^* when $\rho_1^*(\eta_1)$ equals either 0, r or $\frac{\alpha(P_1-P_2)-\beta_1 P_1^2+\beta_2 P_2^2-(\beta_{12}-\beta_{21})P_1P_2-(P_1-P_2)\beta_C\frac{\eta_1}{2}}{\beta_F(P_1-P_2)}$ in the following case A, B and C, respectively.

Case A.

From Lemma 5 (case A), we have $\rho_1^*(\eta_1) = 0$ when

$$P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_C\tfrac{\eta_1}{2}) - P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_C\tfrac{\eta_1}{2}) \leq 0,$$

or equivalently, $\eta_1 \geq \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{\beta_C(P_1-P_2)}$. From Case I, the objective function is independent with η_1 when $\rho_1^*(\eta_1) = 0$. Without loss of generality, when $P_1(\alpha-\beta_1P_1+\beta_{21}P_2) - P_2(\alpha-\beta_2P_2+\beta_{12}P_1) < 0$, we have $\rho_1^*(\eta_1) = 0$ and $\eta_1^* = 0$.

<u>Case B.</u> From Lemma 5 (case B), when $\rho_1^*(\eta_1) = r$, it means

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F r - \beta_C \frac{\eta_1}{2}) \geq P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F r - \beta_C \frac{\eta_1}{2}), \quad (45)$$

or equivalently,

$$\eta_1 \leq \frac{2[P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F r)]}{\beta_C (P_1 - P_2)}. \quad (46)$$

From the proof of Case II, when $\rho_1^*(\eta_1) = r$, $Rev_{n=2}^{2D}(\rho_1^*(\eta_1), \eta_1)$ is concave in η_1. So, we can find the optimal value of η_1 by setting $\frac{\partial Rev_{n=2}^{2D}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1}$ equal to zero. Since

$$\begin{aligned}
\frac{\partial Rev_{n=2}^{2D}(\rho_1^*(\eta_1), \eta_1)}{\partial \eta_1} &= 2r[P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F \frac{r}{2} - \beta_C \eta_1) \\
&\quad - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F \frac{r}{2} - \beta_C \eta_1)],
\end{aligned}$$

$$\frac{\partial Rev_{n=2}^{2D}(\rho_1^*(\eta_1), \eta_1)}{\partial \eta_1} = 0$$

$$\leftrightarrow \eta_1 = \frac{P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F \frac{r}{2}) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F \frac{r}{2})}{\beta_C (P_1 - P_2)}.$$

From the condition (45), since $\eta_1 \in [0, c]$, we have that $P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F \frac{r}{2}) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F \frac{r}{2}) \geq 0$ and $P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F r) \geq 0$. Considering the upper bound on η_1, the optimal η_1^* can be written as

$$\begin{aligned}
\eta_1^* &= \min\{\frac{P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F \frac{r}{2}) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F \frac{r}{2})}{\beta_C (P_1 - P_2)}, \\
&\quad \frac{2[P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F r)]}{\beta_C (P_1 - P_2)}, c\}. \quad (47)
\end{aligned}$$

The following cases identify values of η_1^* and ρ_1^* under different conditions.

Case B.1) If

$$\begin{aligned}
P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) &\geq (P_1 - P_2)(\beta_F \frac{r}{2} + \beta_C c), and \\
P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) &\geq (P_1 - P_2)(\beta_F r + \beta_C \frac{c}{2}),
\end{aligned}$$

we have $\eta_1^* = c$ and $\rho_1^* = r$.

Case B.2) If

$$(P_1-P_2)\beta_F\frac{3r}{2} \le P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1) \le (P_1-P_2)(\beta_F\frac{r}{2}-\beta_Cc),$$

we have $\eta_1^* = \frac{P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_F\frac{r}{2})-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_F\frac{r}{2})}{\beta_C(P_1-P_2)}$ and $\rho_1^*=r$.

Case B.3) If

$$(P_1-P_2)\beta_Fr \le P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1) \le (P_1-P_2)\beta_F\frac{3r}{2}, and$$

$$P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1) \le (P_1-P_2)(\beta_Fr+\beta_C\frac{c}{2})$$

we have $\eta_1^* = \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr)]}{\beta_C(P_1-P_2)}$ and $\rho_1^*=r$.

Case C.

From Lemma 5 (case C), $\rho_1^*(\eta_1) = \frac{\alpha(P_1-P_2)-\beta_1P_1^2+\beta_2P_2^2-(\beta_{12}-\beta_{21})P_1P_2-(P_1-P_2)\beta_C\frac{\eta_1}{2}}{\beta_F(P_1-P_2)}$ if

$$P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_C\tfrac{\eta_1}{2}) \ge P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_C\tfrac{\eta_1}{2}) \text{ and}$$
$$P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr-\beta_C\tfrac{\eta_1}{2}) \le P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr-\beta_C\tfrac{\eta_1}{2})$$

Rearranging the above inequalities, it means

$$\frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr)]}{\beta_C(P_1-P_2)} \le \eta_1,$$

and

$$\eta_1 \le \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{\beta_C(P_1-P_2)}.$$

From the proof of Case III, we have that $Rev_{n=2}^{2D}(\rho_1^*(\eta_1),\eta_1)$ is first concave and then convex in η_1^*. Thus, the optimal η_1 can be obtained by setting the FOC of $Rev_{n=2}^{2D}(\rho_1^*(\eta_1),\eta_1)$ equal 0, where the smaller of the two roots corresponds to the local maximum. Since $\frac{\partial Rev_{n=2}^{2D}(\rho_1^*(\eta_1),\eta_1)}{\partial \eta_1} = 0 \longleftrightarrow$

$\eta_1^* = \{\frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{3\beta_C(P_1-P_2)}, \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{\beta_C(P_1-P_2)}\}$, we have that the local maximum is $\frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{3\beta_C(P_1-P_2)}$.

Therefore, the local optimal η_1^* can be written as,

$$\begin{aligned}\eta_1^* &= \min\{\max\{\frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr)]}{\beta_C(P_1-P_2)},\\ &\quad \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{3\beta_C(P_1-P_2)}\},c\}.\end{aligned}$$

The following cases identify local optimal values of η_1^* and $\rho_1^*(\eta_1)$ under different conditions. **Case C.1)** If

$$(P_1-P_2)\beta_C\frac{3c}{2} \leq P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1) \leq (P_1-P_2)(\beta_C\frac{3c}{2}+\beta_Fr),$$

we have $\eta_1^* = c$ and thus $\rho_1^* = \frac{P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)-(P_1-P_2)\beta_C\frac{c}{2}}{\beta_F(P_1-P_2)}$.

Case C.2) If

$$(P_1-P_2)\beta_F\frac{3r}{2} \leq P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1) \leq (P_1-P_2)\beta_C\frac{3c}{2}, and$$

$$P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1) \leq (P_1-P_2)\beta_F\frac{3r}{2},$$

we have $\eta_1^* = \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{3\beta_C(P_1-P_2)}$ and thus
$\rho_1^* = \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{3\beta_F(P_1-P_2)}$.

Case C.3) If

$$(P_1-P_2)\beta_F\frac{3r}{2} \leq P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1) \leq (P_1-P_2)(\beta_C\frac{c}{2}+\beta_Fr),$$

we have $\eta_1^* = \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr)]}{\beta_C(P_1-P_2)}$ and thus $\rho_1^* = r$.

The solutions identified in each case above represent the local maximum for each feasible region. In some cases, we have multiple local maxima in the same region. In all cases, we found the revenue function is continuous and also continuously differentiable in each region. Next, let us compare the local maximum to determine the global maximum for each of the following cases;

Case 1: $\beta_Cc \leq \beta_Fr$ and $\alpha-\beta_1P_1+\beta_{21}P_2 \leq 1$

when $\beta_Cc \leq \beta_Fr$ and $\alpha-\beta_1P_1+\beta_{21}P_2 \leq 1$, we have the following feasible regions:

Feasible region 1.i) $P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr) \leq 0 \Longrightarrow$ case A applies, where $\rho_1^* = 0$ and $\eta_1^* = 0$.

Feasible region 1.ii) $0 \leq P_1(\alpha-\beta_1P_1+\beta_{21}P_2-\beta_Fr)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1-\beta_Fr) \leq (P_1-P_2)\beta_Fr \Longrightarrow$ case C.2 applies, where $\eta_1^* = \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{3\beta_C(P_1-P_2)}$ and
$\rho_1^* = \frac{2[P_1(\alpha-\beta_1P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2P_2+\beta_{12}P_1)]}{3\beta_F(P_1-P_2)}$.

Feasible region 1.iii) $(P_1 - P_2)\beta_F r \leq P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F r) \leq (P_1 - P_2)\beta_C \frac{3c}{2} \Longrightarrow$ case B.3 and C.2 apply. However, since $Rev^{2D}_{n=2}(\rho_1^*, \eta_1^*)$ of case C.2 is greater than that of case B.3 in this feasible region, the optimal solutions are from case C.2, where $\eta_1^* = \frac{2[P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)]}{3\beta_C(P_1-P_2)}$ and

$\rho_1^* = \frac{2[P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)]}{3\beta_F(P_1-P_2)}$.

Feasible region 1.iv) $(P_1 - P_2)\beta_C \frac{3c}{2} \leq P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F r) \leq (P_1 - P_2)(\beta_F r + \beta_C \frac{c}{2}) \Longrightarrow$ case B.3 and C.1 apply. However, since $Rev^{2D}_{n=2}(\rho_1^*, \eta_1^*)$ of case C.1 is greater than that of case B.3 in this feasible region, the optimal solutions are from case C.1, where $\eta_1^* = c$ and

$\rho_1^* = \frac{P_1(a-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)-(P_1-P_2)\beta_C \frac{c}{2}}{\beta_F(P_1-P_2)}$.

Feasible region 1.v) $(P_1 - P_2)(\beta_F r + \beta_C \frac{c}{2}) \leq P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1 - \beta_F r) \Longrightarrow$ case B.1 applies and thus the optimal solutions are $\eta_1^* = c$ and $\rho_1^* = r$.

Thus, for **Case 1**, where $\beta_C c \leq \beta_F r$, and $\alpha - \beta_1 P_1 + \beta_{21} P_2 \leq 1$, the optimal ρ_1^* and η_1^* can be summarized as:

1.1) $\rho_1^* = 0$ and $\eta_1^* = 0$ when

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) < 0.$$

1.2) $\rho_1^* = r$ and $\eta_1^* = c$ when

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) > (P_1 - P_2)(\beta_F r + \beta_C \tfrac{c}{2}).$$

1.3)

i) $\rho_1^* = \frac{2}{3}[\frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)}{\beta_F(P_1-P_2)}]$ and $\eta_1^* = \frac{2}{3}[\frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)}{\beta_C(P_1-P_2)}]$ when

$$0 \leq P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) \;\; \leq \;\; (P_1 - P_2)\beta_C \frac{3c}{2}.$$

ii) $\rho_1^* = \frac{P_1(\alpha-\beta_1 P_1+\beta_{21})-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)-(P_1-P_2)(\beta_C \frac{c}{2})}{\beta_F(P_1-P_2)}$ and $\eta_1^* = c$ when

$$\begin{aligned}(P_1 - P_2)\beta_C \frac{3c}{2} \quad &< \quad P_1(\alpha - \beta_1 P_1 + \beta_{21}P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12}P_1) \\ &\leq (P_1 - P_2)(\beta_F r + \beta_C \frac{c}{2}).\end{aligned}$$

<u>**Case 2:** $\beta_C c \geq \beta_F r$ **and** $\alpha - \beta_1 P_1 + \beta_{21} P_2 \leq 1$</u>

When $\beta_C c \geq \beta_F r$ and $\alpha - \beta_1 P_1 + \beta_{21} P_2 \leq 1$, we have the following feasible regions:

Feasible Region 2.i) $P_1(\alpha - \beta_1 P_1 + \beta_{21}P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12}P_1 - \beta_F r) \leq 0 \Longrightarrow$ case A applies, where $\rho_1^* = 0$ and $\eta_1^* = 0$.

Feasible region 2.ii) $0 \leq P_1(\alpha-\beta_1 P_1+\beta_{21}P_2-\beta_F r)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1-\beta_F r) \leq (P_1 - P_2)\beta_F r \Longrightarrow$ case C.2 applies, where $\eta_1^* = \frac{2[P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)]}{3\beta_C(P_1-P_2)}$ and $\rho_1^* = \frac{2[P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)]}{3\beta_F(P_1-P_2)}$.

Feasible region 2.iii) $(P_1 - P_2)\beta_F r \leq P_1(\alpha - \beta_1 P_1 + \beta_{21}P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12}P_1 - \beta_F r) \leq (P_1 - P_2)\beta_C \frac{3r}{2} \Longrightarrow$ case B.3 and C.2 apply. However, since $Rev^{2D}_{n=2}(\rho_1^*, \eta_1^*)$ of case C.2 is greater than that of case B.3 in this feasible region, the optimal solutions are from case C.2, where $\eta_1^* = \frac{2[P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)]}{3\beta_C(P_1-P_2)}$ and

$\rho_1^* = \frac{2[P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)]}{3\beta_F(P_1-P_2)}$.

Feasible region 2.iv) $(P_1-P_2)\beta_C \frac{3r}{2} \leq P_1(\alpha-\beta_1 P_1+\beta_{21}P_2-\beta_F r)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1 - \beta_F r) \leq (P_1 - P_2)(\beta_F r + \beta_C \frac{c}{2}) \Longrightarrow$ case B.2 and B.3 apply. However, since $Rev^{2D}_{n=2}(\rho_1^*, \eta_1^*)$ of case B.2 is greater than that of case B.3 in this feasible region, the optimal solutions are from case B.2, where $\eta_1^* = \frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2-\beta_F \frac{r}{2})-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1-\beta_F \frac{r}{2})}{\beta_C(P_1-P_2)}$ and $\rho_1^* = r$.

Feasible region 2.v) $(P_1 - P_2)(\beta_F r + \beta_C \frac{c}{2}) \leq P_1(\alpha - \beta_1 P_1 + \beta_{21}P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12}P_1 - \beta_F r) \leq (P_1 - P_2)(\beta_F \frac{r}{2} + \beta_C c) \Longrightarrow$ case B.2 applies and thus the optimal solutions are $\eta_1^* = \frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2-\beta_F \frac{r}{2})-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1-\beta_F \frac{r}{2})}{\beta_C(P_1-P_2)}$ and $\rho_1^* = r$.

Feasible region 2.vi) $(P_1 - P_2)(\beta_F \frac{r}{2} + \beta_C c) \leq P_1(\alpha - \beta_1 P_1 + \beta_{21}P_2 - \beta_F r) - P_2(\alpha - \beta_2 P_2 + \beta_{12}P_1 - \beta_F r) \Longrightarrow$ B.1 applies and thus the optimal solutions are $\eta_1^* = c$

and $\rho_1^* = r$.

Thus, for **Case 2.**, where $\beta_C c \leq \beta_F r$, and $\alpha - \beta_1 P_1 + \beta_{21} P_2 \leq 1$, the optimal ρ_1^* and η_1^* can be summarized as:

2.1) $\rho_1^* = 0$ and $\eta_1^* = 0$ when

$$P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) > P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2).$$

2.2) $\rho_1^* = r$ and $\eta_1^* = c$

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) > (P_1 - P_2)(\beta_F \tfrac{r}{2} + \beta_C c).$$

2.3)

i) $\rho_1^* = \frac{2}{3}[\frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)}{\beta_F(P_1-P_2)}]$ and $\eta_1^* = \frac{2}{3}[\frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)}{\beta_C(P_1-P_2)}]$ when

$$0 \leq P_1(\alpha - \beta_1 P_1 + \beta_{21}) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) \quad \leq \quad (P_1 - P_2)\beta_F \frac{3r}{2}.$$

ii) $\rho_1^* = r$ and $\eta_1^* = \frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)-(P_1-P_2)\beta_F \frac{r}{2}}{\beta_C(P_1-P_2)}$ when

$$\begin{aligned}(P_1 - P_2)\beta_F \frac{3r}{2} \quad &< \quad P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) \\ &\qquad - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) \leq (P_1 - P_2)(\beta_F \frac{r}{2} + \beta_C c).\end{aligned}$$

In the case where $\alpha - \beta_1 P_1 + \beta_{21} P_2 > 1$, there is more demand (at price P_1) than the capacity for the front-center section. Consider the constraint (43): $\alpha - \beta_1 P_1 + \beta_{21} P_2 - \beta_F \frac{\rho_1}{2} - \beta_C \frac{\eta_1}{2} \leq 1$, which is equivalent to $\beta_F \rho_1 + \beta_C \eta_1 \geq 2(\alpha - \beta_1 P_1 + \beta_{21} P_2 - 1) > 0$. From Assumption 5, it can be written as $\rho_1 \geq M_F = \frac{\alpha-\beta_1 P_1+\beta_{21}P_2-1}{\beta_F}$ and $\eta_1 \geq M_C = \frac{\alpha-\beta_1 P_1+\beta_{21}P_2-1}{\beta_C}$. The proof of this case is similar to the proofs above (where $\alpha - \beta_1 P_1 + \beta_{21} P_2 \leq 1$). The only difference is that now we are solving the optimization problem over the interval $[M_F, r]$ for the decision variable ρ_1 and $[M_C, c]$ for the decision variable η_1. The optimal solutions when $\alpha - \beta_1 P_1 + \beta_{21} P_2 > 1$ can be summarized as followed;

Case 3: $\alpha - \beta_1 P_1 + \beta_{21} P_2 > 1$ **and** $\beta_C c \leq \beta_F r$

when $\alpha - \beta_1 P_1 + \beta_{21} P_2 > 1$ and $\beta_C c \leq \beta_F r$, the optimal decisions for two-dimensional seating sections are given by;

3.1) Offer all seats with high price P_1 ($\rho_1^* = r$ and $\eta_1^* = c$) when

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) > (P_1 - P_2)(\beta_F r + \beta_C \tfrac{c}{2}).$$

3.2) Offer two seating sections with P_1 and P_2 where:

i) $\rho_1^* = M_F$ and $\eta_1^* = M_C$ when

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) > (P_1 - P_2)(\beta_F M_F + \beta_C M_C).$$

ii) $\rho_1^* = \frac{2}{3}[\frac{P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1)}{\beta_F (P_1 - P_2)}]$ and $\eta_1^* = \frac{2}{3}[\frac{P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1)}{\beta_C (P_1 - P_2)}]$ when

$$\begin{aligned}(P_1 - P_2)(\beta_F M_F + \beta_C M_C) \leq\ & P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) \\ & - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) \leq (P_1 - P_2)\beta_C \frac{3c}{2}\end{aligned} \quad (48)$$

iii) $\rho_1^* = \frac{P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) - (P_1 - P_2)(\beta_C \frac{c}{2})}{\beta_F (P_1 - P_2)}$ and $\eta_1^* = c$ when

$$\begin{aligned}(P_1 - P_2)\beta_C \frac{3c}{2} <\ & P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) \\ & - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) \leq (P_1 - P_2)(\beta_F r + \beta_C \frac{c}{2}).\end{aligned}$$

Case 4: $\alpha - \beta_1 P_1 + \beta_{21} P_2 > 1$ **and** $\beta_C c > \beta_F r$

When $\alpha - \beta_1 P_1 + \beta_{21} P_2 > 1$ and $\beta_C c > \beta_F r$, the optimal decisions for two-dimensional seating sections are given by:

4.1) Offer all seats with high price P_1 ($\rho_1^* = r$ and $\eta_1^* = c$) when

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) > (P_1 - P_2)(\beta_F \tfrac{r}{2} + \beta_C c).$$

4.2) Offer two seating sections with P_1 and P_2 where:

i) $\rho_1^* = M_F$ and $\eta_1^* = M_C$ when

$$P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) > (P_1 - P_2)(\beta_F M_F + \beta_C M_C).$$

ii) $\rho_1^* = \frac{2}{3}[\frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)}{\beta_F(P_1-P_2)}]$ and $\eta_1^* = \frac{2}{3}[\frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)}{\beta_C(P_1-P_2)}]$ when

$$(P_1 - P_2)(\beta_F M_F + \beta_C M_C) < P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) \leq (P_1 - P_2)\beta_F \tfrac{3r}{2}.$$

iii) $\rho_1^* = r$ and $\eta_1^* = \frac{P_1(\alpha-\beta_1 P_1+\beta_{21}P_2)-P_2(\alpha-\beta_2 P_2+\beta_{12}P_1)-(P_1-P_2)\beta_F \frac{r}{2}}{\beta_C(P_1-P_2)}$ when

$$\begin{aligned}(P_1 - P_2)\beta_F \frac{3r}{2} &< P_1(\alpha - \beta_1 P_1 + \beta_{21} P_2) \\ &\quad - P_2(\alpha - \beta_2 P_2 + \beta_{12} P_1) \leq (P_1 - P_2)(\beta_F \frac{r}{2} + \beta_C c).\end{aligned}$$

We summarize the optimal zoning decisions for all cases (1, 2, 3 and 4) as written in Theorem 4. ■

Proof for Lemma 6

Proof. The revenue maximization problem given in (20) subject to (21) can be written as

$$\max_{r_1,\ldots,r_{n-1}\in[0,r]} Rev^{1D} = \sum_{i=1}^{n} P_i(r_i - r_{i-1})(a - b_i P_i + \sum_{j\neq i} b_{ji} P_j - b_F \frac{r_i + r_{i-1}}{2}),$$

subject to:

$$a - b_i P_i + \sum_{j\neq i} b_{ji} P_j - b_F \frac{r_i + r_{i-1}}{2} \quad \leq \quad row_c, \qquad i \in \{1, \ldots, n\},$$

where $r_0 = 0$ and $r_n = r$. Because

$$\begin{aligned}
\frac{\partial^2 Rev^{1D}}{\partial r_i^2} &= -b_F(P_i - P_{i+1}), i \in \{1, \ldots, n-1\}, \\
\frac{\partial^2 Rev^{1D}}{\partial r_i \partial r_j} &= 0, j \neq i,
\end{aligned}$$

the hessian of the objective function is as follow:

$$H_{Rev^{1D}} = \begin{pmatrix} -b_F(P_1 - P_2) & 0 & \cdots & 0 \\ 0 & -b_F(P_2 - P_3) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -b_F(P_{n-1} - P_n) \end{pmatrix},$$

whose leading coefficient is negative and the determinants alternate in sign. Thus, the hessian is negative definite and the objective function is jointly concave in $r_1, \ldots, r_{n-1}$.

■

Proof for Theorem 5

Proof. After rearranging terms, we can write the revenue maximization problem given in (22) as

$$\max_{0 \le r_1 \le r} Rev_{n=2}^{1D} = P_1 r_1 (a - b_1 P_1 + b_{21} P_2 - b_F \frac{r_1}{2}) + P_2 (r - r_1)(a - b_2 P_2 + b_{12} P_1 - b_F \frac{(r + r_1)}{2}),$$

subject to:

$$\frac{2(a - b_1 P_1 + b_{21} P_2 - row_c)}{b_F} \le r_1, \tag{49}$$

$$\frac{2(a - b_2 P_2 + b_{12} P_1 - row_c)}{b_F} - r \le r_1. \tag{50}$$

Note that from the assumption **A3** (iv), $\mathrm{E}[\int_0^r (a - b_2 P_2 + b_{12} P_1 - b_F D) dD] \le r \times row_c$. So, we have $\frac{2(a - b_2 P_2 + b_{12} P_1 - row_c)}{b_F} - r \le 0$ and the constraint (50) is redundant. As defined earlier, $m = \frac{2(a - b_1 P_1 + b_{21} P_2 - row_c)}{b_F}$. Thus, the problem can be reduced to:

$$Rev_{n=2}^{1D} = \max_{max\{0,m\} \le r_1 \le r} P_1 r_1 (a - b_1 P_1 + b_{21} P_2 - b_F \frac{r_1}{2}) + P_2 (r - r_1)(a - b_2 P_2 + b_{12} P_1 - b_F \frac{(r + r_1)}{2}).$$

From Lemma 6, the objective function is concave in the decision variable, r_1 (when n=2, $\frac{\partial^2 Rev^{1D}_{n=2}}{\partial r_1^2} = b_F(P_2 - P_1) < 0$). Therefore, it allows us to determine the optimal value of r_1 by setting $\frac{\partial Rev^{1D}_{n=2}}{\partial r_1}$ equal 0.

In case 1, $m \leq 0$, then $max\{m, 0\} = 0$. When $P_2(a-b_2P_2+b_{12}P_1) > P_1(a-b_1P_1+b_{21}P_2)$ in case 1 (i), we have that $Rev^{1D}_{n=2}$ is decreasing in the interval $[0, r]$ and $r_1^* = 0$. In case 2, $m > 0$, then $max\{m, 0\} = m$. When $P_2(a - b_2P_2 + b_{12}P_1 - b_Fm) > P_1(a - b_1P_1+b_{21}P_2-b_Fm)$ in case 2 (i), $Rev^{1D}_{n=2}$ is decreasing in the interval $[m, r]$ and $r_1^* = m$. In case (ii) when $P_1(a-b_1P_1+b_{21}P_2-b_Fr) > P_2(a-b_2P_2+b_{12}P_1-b_Fr)$, we have that $Rev^{1D}_{n=2}$ is increasing in the interval $[max\{m, 0\}, r]$ and $r_1^* = r$. Otherwise, the revenue function is increasing and then decreasing and $r^* = \frac{a(P_1-P_2)-b_1P_1^2+b_2P_2^2-P_1P_2(b_{12}-b_{21})}{b_F(P_1-P_2)}$, which is equivalent to $P_1(a - b_1P_1 + b_{21}P_2 - b_Fr_1^*) = P_2(a - b_2P_2 + b_{12}P_1 - b_Fr_1^*)$. ■

Proof for Lemma 7

Proof. Suppose $0 < r_1^* < 0$, then $r_1^* = \frac{a(P_1-P_2)-b_1P_1^2+b_2P_2^2-P_1P_2(b_{12}-b_{21})}{b_F(P_1-P_2)}$. Since $P_1 > P_2$, we have $\frac{\partial r_1^*}{\partial b_1} = -\frac{P_1^2}{b_F(P_1-P_2)} < 0$. From Theorem 5 case 1(i), 2(i) where $r_1^* = 0$ and $r_1^* = m$, the conditions for 0 and m to be optimal are easier to satisfy as b_1 increases. For case 1(ii) and 2(ii), the condition is more difficult to satisfy to have $r_1^* = r$. Thus, as b_1 increases, the optimal zoning decision is moving away from r and toward 0 and m; that is the optimal row, r_1^*, should be moved to be closer to the stage.

Similar arguments hold for b_2 comparative statics, and we have $\frac{\partial r_1^*}{\partial b2} = \frac{P_2^2}{b_F(P_1-P_2)} > 0$. From Theorem 5, we can see that condition for $r_1^* = r$ to be optimal is easier to satisfy as b_2 increases; so the optimal zoning decision is moving toward r and away from 0 and m. Therefore, the optimal row, r_1^*, should be moved to be further from the stage as b_2 increases. ■

Proof for Lemma 8

Proof. From Theorem 5 case (iii), if $r_1^* \in (0, r)$, $r_1^* = \frac{a(P_1-P_2)-b_1P_1^2+b_2P_2^2-P_1P_2(b_{12}-b_{21})}{b_F(P_1-P_2)}$. Taking the partial derivative of the optimal solution with respect to b_{12} and b_{21}, we have:

$$\frac{\partial r_1^*}{\partial b_{12}} = -\frac{P_1P_2}{b_F(P_1-P_2)} < 0 \text{ and } \frac{\partial r_1^*}{\partial b_{21}} = \frac{P_1P_2}{b_F(P_1-P_2)} > 0.$$

Thus, r_1^* is decreasing in b_{12} and increasing in b_{21}. In the conditions for r_1^* to equal 0, m, or r, we can see that as b_{12} increases (and as b_{21} decreases), the optimal zoning decision is moving away from r and toward 0 and m, resulting 0 and m to be easier to satisfy. Thus, as the marginal price substitution effect b_{12} (b_{21}) increases, the optimal row should be moved to be closer to (further from) the stage. ■

Proof for Lemma 9

Proof. Suppose $r_1^* \in (0, r)$, then $r_1^* = \frac{a(P_1-P_2)-b_1P_1^2+b_2P_2^2-P_1P_2(b_{12}-b_{21})}{b_F(P_1-P_2)}$.. Taking the partial derivative of the optimal solution with respect to b_F, we have

$$\frac{\partial r_1^*}{\partial b_F} = -\frac{a(P_1-P_2)-b_1P_1^2+b_2P_2^2-P_1P_2(b_{12}-b_{21})}{(b_F)^2(P_1-P_2)} = -\frac{r_1^*}{b_F} \leq 0.$$

Note that the condition for $r_1^* = 0$ (from Theorem 5 case 1 (i)) is not sensitive to b_F. The condition for $r_1^* = m$ from case 2(i) is $P_2(a - b_2P_2 + b_{12}P_1 - b_Fm) > P_1(a-b_1P_1+b_{21}P_2-b_Fm)$ or $P_2(a-b_2P_2+b_{12}P_1) > P_1(a-b_1P_1+b_{21}P_2)-(P_1-P_2)b_Fm$. We can see that as b_F increases, this condition is easier to to satisfy. Likewise, as b_F increases, the condition for $r_1^* = r$ is more difficult to satisfy. So, the optimal zoning decision is moving away from r; i.e., the optimal row, r_1^*, should be moved to be closer to the stage as b_F increases. ■

Comparative statics of the optimal zoning decisions and optimal revenue shown in Table 8

When $r_1^* \in (0, r)$, we have that $r_1^* = \frac{a(P_1-P_2)-b_1P_1^2+b_2P_2^2-P_1P_2(b_{12}-b_{21})}{b_F(P_1-P_2)}$. The optimal revenue is

$$Rev_{n=2}^{1D}(r_1^*) = P_1r_1^*(a-b_1P_1+b_{21}P_2-b_F\frac{(r_1^*)^2}{2})+P_2(r-r_1^*)(a-b_2P_2+b_{12}P_1-b_F\frac{(r+r_1^*)^2}{2}).$$

Taking the derivatives of the optimal decision and the optimal revenue with respect to each model parameter, we summarize the results in Table 9 below.

Table 9: Comparative statics on the optimal zoning decision and the optimal revenue.

	$\partial r_1^* /$	$\partial Rev_{n=2}^{*1D} /$
∂a	$\frac{1}{b_F} \geq 0$	$(P_1 - P_2) r_1^* + r P_2 \geq 0$
∂b_1	$-\frac{P_1^2}{b_F(P_1 - P_2)} \leq 0$	$-r_1^*(P_1)^2 \leq 0$
∂b_2	$\frac{P_2^2}{b_F(P_1 - P_2)} \geq 0$	$(P_2)^2(r_1^* - r) \leq 0$
∂b_{12}	$\frac{P_1 P_2}{b_F(P_1 - P_2)} \leq 0$	$P_1 P_2 (r - r_1^*) \geq 0$
∂b_{21}	$\frac{P_1 P_2}{b_F(P_1 - P_2)} \geq 0$	$P_1 P_2 r_1^* \geq 0$
∂b_F	$-\frac{r_1^*}{b_F} \leq 0$	$-\frac{(P_1 - P_2)(r_1^*)^2 - P_2 r^2}{2} \leq 0$
∂P_1	$-\frac{(b_2 + b_{21} - b_{12})(P_2)^2 - 2b_1 P_1 P_2 + b_1 (P_1)^2}{b_F(P_1 - P_2)}$	$\frac{2r_1^*(a - 2b_1 P_1 - b_{12} P_2 - b_{21} P_2) - b_F (r_1^*)^2 + 2b_{12} P_2 r}{2}$
∂P_2	$-\frac{(b_1 + b_{12} - b_{21})(P_1)^2 - 2b_2 P_1 P_2 + b_2 (P_2)^2}{b_F(P_1 - P_2)}$	$\frac{2(r - r_1^*)(a + b_{21} P_1 - 2b_2 P_2) - b_F(r^2 - (r_1^*)^2) + 2b_{21} P_1 r_1^*}{2}$

Proof for Lemma 10

Proof.

i) When $r_1^* \in (0, r)$, we have $r_1^* = \frac{a(P_1 - P_2) - b_1 P_1^2 + b_2 P_2^2 - P_1 P_2 (b_{12} - b_{21})}{b_F(P_1 - P_2)}$. Considering how r_1^* changes as P_1 changes, we take the partial derivative of the optimal solution with respect to P_1:

$$\frac{\partial r_1^*}{\partial P_1} = -\frac{(b_2 + b_{21} - b_{12})(P_2)^2 - 2b_1 P_1 P_2 + b_1 (P_1)^2}{b_F(P_1 - P_2)}.$$

Rearranging the equation above, we have $\frac{\partial r_1^*}{\partial P_1} < 0$ if $b_1(P_1)^2 + b_2(P_2)^2 + b_{21}(P_2)^2 > 2b_1 P_1 P_2 + b_{12}(P_2)^2$, and $\frac{\partial r_1^*}{\partial P_1} > 0$, otherwise.

ii) Similarly for P_2, we have

$$\frac{\partial r_1^*}{\partial P_2} = -\frac{(b_1 + b_{12} - b_{21})(P_1)^2 - 2b_2 P_1 P_2 + b_2 (P_2)^2}{b_F(P_1 - P_2)}.$$

Rearranging the equation above, we have $\frac{\partial r_1^*}{\partial P_2} < 0$ if $b_1(P_1)^2 + b_2(P_2)^2 + b_{12}(P_1)^2 > 2b_2 P_1 P_2 + b_{21}(P_1)^2$, and $\frac{\partial r_1^*}{\partial P_1} > 0$, otherwise. ■

APPENDIX C

APPENDIX FOR CHAPTER 4

Proof for Theorem 6

Proof. The results in this Proposition can be proved by induction on t. Consider following two induction hypotheses:

i) $f(\gamma|p_t, p_{t+1}, \ldots, p_n, m_{t+1}, \ldots, m_n)$
$= \frac{e^{-\gamma(b+\sum_{k=t+1}^{n}\phi(p_n)g(n))}\gamma^{a+\sum_{k=t+1}^{n}m_k}[b+\sum_{k=t+1}^{n}\phi(p_n)g(n)]^{a+\sum_{k=t+1}^{n}m_k}}{(a+\sum_{k=t+1}^{n}m_k-1)!}, \forall\gamma > 0.$

That is the distribution function of the base demand rate in period t, Γ_t, which follows a Gamma distribution with a scale parameter of $a + \sum_{k=t+1}^{n} m_k$ and a shape parameter of $b + \sum_{k=t+1}^{n} \phi(p_n)g(n)$.

ii) $f(M_t = m|p_t, p_{t+1}, \ldots, p_n, m_{t+1}, \ldots, m_n)$
$= \binom{m+a+\sum_{k=t+1}^{n}m_k-1}{m}\left(\frac{\phi(p_t)g(t)}{A(t)}\right)^m\left(\frac{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)}{A(t)}\right)^{a+\sum_{k=t+1}^{n}m_k}, \forall m \in \{0, 1, 2, \ldots\},$
where $A(t) = b + \sum_{k=t+1}^{n}\phi(p_k)g(k) + \phi(p_t)g(t)$.

That is the distribution function of the unconditional ticket demand in period t, M_t, which follows a Negative Binomial distribution with parameters $a+\sum_{k=t+1}^{n} m_k$ and
$\frac{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)}{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)+\phi(p_t)g(t)}$.

If $t = n$, the hypothesis (i) holds by the model definition that $\Gamma_{t=n}$, follows a Gamma distribution with a scale parameter of a and a shape parameter of b. In addition, the hypothesis (ii) also holds when $t = n$ since from equation (27), we have that $f(M_t = m|p_t)$ follows a Negative Binomial distribution with parameters a and $\frac{b}{b+\phi(p_t)g(t)}$.

Assume that the induction hypotheses (i) and (ii) are true for $t = j + 1$, and consider the case when $t = j$. If (i) and (ii) hold for $t = j + 1$, we have:

A.1) $f(\Gamma_{j+1} = \gamma | p_{j+1}, p_{j+2}, \ldots, p_n, m_{j+2}, \ldots, m_n)$
$= \frac{e^{-\gamma(b+\sum_{k=j+2}^{n} \phi(p_n)g(n))} \gamma^{a+\sum_{k=j+2}^{n} m_k} [b+\sum_{k=j+2}^{n} \phi(p_n)g(n)]^{a+\sum_{k=j+2}^{n} m_k}}{(a+\sum_{k=j+2}^{n} m_k - 1)!}, \forall \gamma > 0;$

B.1) $f(M_{j+1} = m | p_{j+1}, p_{j+2}, \ldots, p_n, m_{j+2}, \ldots, m_n)$
$= \binom{m+a+\sum_{k=j+2}^{n} m_k - 1}{m} \left(\frac{\phi(p_{j+1})g(j+1)}{A(j+1)}\right)^m \left(\frac{b+\sum_{k=j+2}^{n} \phi(p_k)g(k)}{A(j+1)}\right)^{a+\sum_{k=j+2}^{n} m_k},$
$\forall m \in \{0, 1, 2, \ldots\},$
where $A(j+1) = b + \sum_{k=j+2}^{n} \phi(p_k)g(k) + \phi(p_{j+1})g(j+1)$.

Then, consider the case where $t = j$;

A.2) Using Bayes' rule, we have that the posterior distribution of Γ_j is equal to:

$$\begin{aligned}
&f(\Gamma_j = \gamma | p_j, p_{j+1}, \ldots, p_n, m_{j+1}, \ldots, m_n) \\
&\quad = \frac{f(M_{j+1} = m | p_{j+1}, p_{j+2}, \ldots, p_n, m_{j+2}, \ldots, m_n) f(\gamma)}{\int_0^\infty f(M_{j+1} = m | p_{j+1}, p_{j+2}, \ldots, p_n, m_{j+2}, \ldots, m_n) f(\gamma) d\gamma} \\
&\quad = \frac{e^{-\gamma(b+\sum_{k=j+1}^{n} \phi(p_n)g(n))} \gamma^{a+\sum_{k=j+1}^{n} m_k} [b + \sum_{k=j+1}^{n} \phi(p_n)g(n)]^{a+\sum_{k=j+1}^{n} m_k}}{(a + \sum_{k=j+1}^{n} m_k - 1)!}, \forall \gamma > 0,
\end{aligned}$$

where $f(M_{j+1} = m | p_{j+1}, p_{j+2}, \ldots, p_n, m_{j+2}, \ldots, m_n)$ is given by (B.1).

B.2) The distribution of demand in period $t = j$, unconditional on Γ_j, is given by:

$$\begin{aligned}
&f(M_j = m | p_j, p_{j+1}, \ldots, p_n, m_{j+1}, \ldots, m_n) \\
&\quad = \int_0^\infty f(M_j = m | \Gamma_j = \gamma, p_j, \ldots, p_n, m_{j+1}, \ldots, m_n) f(\gamma | p_j, \ldots, p_n, m_{j+1}, \ldots, m_n) d\gamma \\
&\quad = \int_0^\infty \left(\frac{[\phi(p_j)g(j)\gamma]^m e^{-\phi(p_j)g(j)\gamma}}{m!}\right) f(\gamma | p_j, \ldots, p_n, m_{j+1}, \ldots, m_n) d\gamma \\
&\quad = \binom{m + a + \sum_{k=j+1}^{n} m_k - 1}{m} \left(\frac{\phi(p_j)g(j)}{A(j)}\right)^m \left(\frac{b + \sum_{k=t+1}^{n} \phi(p_k)g(k)}{A(j)}\right)^{a+\sum_{k=j+1}^{n} m_k}, \\
&\quad \forall m \in \{0, 1, 2, \ldots\},
\end{aligned}$$

where $A(j) = b + \sum_{k=j+1}^{n} \phi(p_k)g(k) + \phi(p_j)g(j)$, and $f(\gamma | p_t, p_{j+1}, \ldots, p_n, m_{j+1}, \ldots, m_n)$ is given by (A.2).

From (A.2) and (B.2), it follows that the induction hypothesis (i) and (ii) are true for $t = j$, given they are true for $t = j + 1$. Thus, the results in this Proposition follow by induction.

Proof for Theorem 7

Proof. Let $mean_{dt}$ denote the expected mean of demand in period t. Since the posterior distribution of demand is Negative Binomial with parameters $a + \sum_{k=t+1}^{n} m_k$ and $\frac{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)}{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)+\phi(p_t)g(t)}$, the expected mean of demand ($mean_{dt}$) equals $\frac{[a+\sum_{k=t+1}^{n} m_k]\phi(p_t)g(t)}{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)}$. The proofs of results given in Proposition 7 are as follow;

i) The derivative of $mean_{dt}$ with respect to the ticket price, p_t, is:

$$\frac{\partial mean_{dt}}{\partial p_t} = \frac{(a+\sum_{k=t+1}^{n} m_k)g(t)}{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)}\Big[\frac{\partial\phi(p_t)}{\partial p_t}\Big].$$

Since $\frac{(a+\sum_{k=t+1}^{n} m_k)g(t)}{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)} > 0$ and $\frac{\partial\phi(p_t)}{\partial p_t} < 0$, we have $\frac{\partial mean_{dt}}{\partial p_t} < 0$. Therefore, the expected demand in period t is decreasing with the ticket price in period t.

ii) Taking the first derivative of $mean_{dt}$ with respect to prices used in the past periods, p_i, $i = t+1, \ldots, n$, we have:

$$\frac{\partial mean_{dt}}{\partial p_i} = \frac{-(a+\sum_{k=t+1}^{n} m_k)\phi(p_t)g(t)}{[b+\sum_{k=t+1}^{n}\phi(p_k)g(k)]^2}\Big(\frac{\partial\phi(p_i)}{\partial p_i}\Big)g(i), \quad \forall i = t+1, \ldots, n.$$

Since $\frac{(a+\sum_{k=t+1}^{n} m_k)\phi(p_t)g(t)}{[b+\sum_{k=t+1}^{n}\phi(p_k)g(k)]^2} > 0$ and $\frac{\partial\phi(p_i)}{\partial p_i}g(i) < 0, \forall i = t+1, \ldots, n$, it follows that $\frac{\partial mean_{dt}}{\partial p_i} > 0, \forall i = t+1, \ldots, n$. In other words, the expected demand in period t is decreasing with the selling prices offered in the past periods $t+1, \ldots, n$.

iii) The derivative of $mean_{dt}$ with respect to the observed sales in the past periods (m_i, where $i = t+1, \ldots, n$) is as follows:

$$\frac{\partial mean_{dt}}{\partial m_i} = \frac{\phi(p_t)g(t)}{b+\sum_{k=t+1}^{n}\phi(p_k)g(k)} > 0, \quad \forall i = t+1, \ldots, n.$$

Thus, the expected demand in period t is increasing with sales in the past periods, m_i, where $i = t+1, \ldots, n$.

REFERENCES

[1] Aviv, Y. and Pazgal, A., "Pricing of short life-cycle products through active learning." Working Paper, Washington University, St. Louis, 2002.

[2] Aviv, Y. and Pazgal, A., "A partially observed Markov decision process for dynamic pricing," *Management Science*, vol. 51, p. 14001416, 2005.

[3] Azoury, K., "Bayes solution to dynamic inventory models under unknown demand distribution," *Management Science*, vol. 12, pp. 1150–1160, 1985.

[4] Balvers, R. and Cosimano, T., "Actively learning about demand and the dynamics of price adjustment," *The Economic Journal*, vol. 100, pp. 882–898, 1990.

[5] Baskerville, D., *Music business handbook and career guide*. California: Sage Publications, 2001.

[6] Bassamboo, A., Harrison, J., and Zeevi, A., "Design and control of a large call center: Asymptotic analysis of an LP-based method.," *Operations Research*, vol. 54, no. 3, pp. 419–435, 2006.

[7] Belobaba, P., "Airline yield management: An overview of seat inventory control," *Transportation Science*, vol. 21, pp. 63–73, 1987.

[8] Benjaafar, S., "Performance bounds for the effectiveness of pooling in multi-processing systems," *European Journal of Operational Research*, vol. 87, pp. 375–388, 1995.

[9] Benjaafar, S., Cooper, W., and Kim, J., "On the benefits of pooling in production-inventory systems," *Management Science*, vol. 51, pp. 548–565, 2005.

[10] Best Friends, 2008. Official Website of Best Friends Animal Society, http://www.bestfriends.org, last accessed on January 28, 2008.

[11] Biller, S., Chan, L., Simchi-Levi, D., and Swann, J., "Dynamic pricing and the direct-to-customer model in the automotive industry," *Electronic Commerce Research*, vol. 5, no. 2, pp. 309–334, 2005.

[12] Bitran, G. and Gilbert, S., "Managing hotel reservations with uncertain arrivals," *Operations Research*, vol. 44, pp. 15–49, 1996.

[13] Bitran, G. and Mondschein, S., "An application of yield management to the hotel industry considering multiple stays," *Operations Research*, vol. 43, pp. 427–443, 1995.

[14] Bitran, G. and Wadhwa, H., "A methodology for demand learning with an application to the optimal pricing of seasonal products." Working Paper, M.I.T. Sloan School of Management, 1996.

[15] Bitran, G., Haas, E., and Matsuo, H., "Production planning of style goods with high setup costs and forecast revisions," *Operations Research*, vol. 34, pp. 226–236, 1986.

[16] Bitran, G. and Mondschein, S., "Periodic pricing of seasonal products in retailing," *Management Science*, vol. 43, pp. 64–79, 1997.

[17] Bitran, G. Caldentey, R. and Mondschein, S., "Coordinating clearance markdown sales of seasonal products in retail chains," *Operations Research*, vol. 46, pp. 609–624, 1998.

[18] Bitran, R. and Caldentey, R., "Pricing models for revenue management.," *Manufacturing and Service Operations Management*, vol. 5, no. 3, pp. 203–229, 2003.

[19] Borland, J. and MacDonald, R., "Demand for sport," *Oxford Review of Economic Policy*, vol. 19, no. 4, pp. 478–502, 2003.

[20] Borst, S., Mandelbaum, A., and Reiman, M., "Dimensioning large call centers," *Operations Research*, vol. 52, pp. 17–34, 2004.

[21] Bradford, J. and Sugrue, P., "A Bayesian approach to the two-period style goods inventory problem with single replenishment and heterogeneous Poisson demands," *Journal of the Operational Research Society*, vol. 41, pp. 211–218, 1990.

[22] Brockmeyer, E., Halstrom, H., and Jensen, A., "The life and works of A. K. Erlang," *Danish Academy of Technical Sciences, Copenhagen*, 1948.

[23] Buzacott, J., "Commonalities in reengineered business process: Models and issues," *Management Science*, vol. 42, pp. 768–782, 1996.

[24] Caro, F. and Simchi-Levi, D., "Static pricing for a network service provider." Working Paper, University of California, Los Angeles, Anderson School of Management, 2006.

[25] Carol, W. and Grimes, R., "Evolutionary change in product management: Experiences in the car rental industry," *Interfaces*, vol. 25, pp. 84–104, 1995.

[26] Carrier, E., "Modeling airline passenger choice: Passenger preference for schedule in the Passenger Oriented-Destination Simulator (PODS)," *Master Thesis. Massachusetts Institute of Technology*, 2003.

[27] Centonze, L. and Levy, J., "Characteristics of free-roaming cats and their caretakers," *Journal of the American Veterinary Medical Association*, vol. 220, no. 11, pp. 1627–1633, 2002.

[28] Chan, L., Shen, Z., Simchi-Levi, D., and Swann, J., "Coordination of pricing and inventory decisions: A survey and classification. in: D Simchi-Levi et al. (Eds), Handbook of quantitative supply chain analysis: Modeling in the E-Business Era.," *Kluwer Academic Publishers*, pp. 335–392, 2004.

[29] Chang, S. and Fyfee, D., "Estimation of forecast errors for seasonal style goods sales," *Management Science*, vol. 18, pp. 89–96, 1971.

[30] Chatwin, R., "Optimal dynamic pricing of perishable products with stochastic demand and a finite set of prices," *European Journal of Operational Research*, vol. 125, pp. 149–174, 2000.

[31] Ciancimino, A., Inzerillo, G., Lucidi, S., and Palagi, L., "A mathematical programming approach for the solution of the railway yield management problem," *Transportation Science*, vol. 33, pp. 168–181, 1999.

[32] Cooper, R., *Introduction to Queueing Theory, second edition.* North Holland, New York, 1981.

[33] Corning, J. and Levy, A., "Demand for live theater with market segmentation and seasonality," *Journal of Cultural Economics*, vol. 26, pp. 217–235, 2002.

[34] Courty, P., "An economic guide to ticket pricing in the entertainment industry," *Louvain Economic Review*, vol. 66, no. 1, pp. 167–192, 2000.

[35] Courty, P., "Ticket pricing under demand uncertainty," *Journal of Law and Economics*, vol. 46, no. 2, pp. 627–652, 2003.

[36] Courty, P., "The impact of price discrimination of revenue: Evidence from the concert industry." Working paper, 2009.

[37] DeGraba, P., "Buying frenzies and seller-induced excess demand," *RAND Journal of Economics.*, vol. 26, no. 2, pp. 331–342, 1995.

[38] DeSerpa, A., "To err is rational : A theory of excess demand for tickets," *Managerial and Decision Economics*, vol. 15, no. 5, pp. 511–518, 1994.

[39] Dewan, S. and Mendelson, H., "User delay costs and internal pricing for a service facility," *Management Science*, vol. 36, pp. 1502–1517, 1990.

[40] Diesel, G., Smith, H., and Pfeiffer, D., "Factors affecting time to adoption of dogs re-homed by a charity in the UK," *Animal Welfare*, vol. 16, no. 3, pp. 353–360, 2007.

[41] Drake, M. J., Duran, S., Griffin, P., and Swann, J., "Optimal timing of switches between product sales for sports and entertainment tickets," *Naval Research Logistics*, vol. 55, no. 1, pp. 59–75, 2008.

[42] Einav, L. and Orbach, B., "Uniform prices for differentiated goods: The case of the movie theater industry," *International Review of Law and Economics*, vol. 27, no. 2, pp. 129–153, 2007.

[43] Ekelund, R. J. and Ritenour, S., "An exploration of the Beckerian Theory of Time Costs: Symphony concert demand," *The American Journal of Economics and Sociology*, vol. 58, pp. 887–889, 1999.

[44] Elmaghraby, W. and Keskinocak, P., "Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions," *Management Science.*, vol. 49, pp. 1287–1309, 2003.

[45] Eppen, G. and Iyer, A., "Improved fashion buying with Bayesian updates," *Operations Research*, vol. 45, pp. 805–819, 1997.

[46] Erlang, A., "On the rational determination of the number of circuits, the life and works of A. K. Erlang, H. L. Halstrom and A. Jensen (eds.)," *Danish Academy of Technical Sciences, 1948*, pp. 216–221, 1924.

[47] Felton, M. V., "Major influences on the demand for opera tickets," *Journal of Cultural Economics*, vol. 13, pp. 53–64, 1989.

[48] Feng, Y. and Gallego, G., "Optimal starting times for end-of-season sales and optimal stopping times for promotional fares," *Management Science*, no. 41.

[49] Feng, Y. and Xiao, B., "Maximizing revenues of perishable assets with a risk factor," *Operations Research*, vol. 47, pp. 337–341, 1999.

[50] Fisher, M. and Raman, A., "Reducing the cost of demand uncertainty through accurate response to early sales," *Operations Research*, vol. 44, pp. 87–99, 1996.

[51] Gallego, G. and Van Ryzin, G., "Optimal dynamic pricing of inventories with stochastic demand over finite horizons," *Management Science*, vol. 40, no. 8, pp. 999–1020, 1994.

[52] Gallego, G. and VanRyzin, G., "A multi-product dynamic pricing problem and its applications to network yield management," *Operations Research*, vol. 45, pp. 24–41, 1997.

[53] Gans, N. and Zhou, Y., "A call-routing problem with service-level constraints," *Operations Research*, vol. 51, no. 2, pp. 255–271, 2003.

[54] Garnett, O., Mandelbaum, A., and Reiman, M., "Designing a call center with impatient customers," *Manufacturing and Service Operations Management*, vol. 4, no. 3, pp. 208–227, 2002.

[55] Geraghty, M. K. and Johnson, E., "Revenue management saves National car rental," *Interfaces*, vol. 27, no. 1, pp. 107–127, 1997.

[56] Gil, R. and Hartmann, W. R., "Empirical analysis of metering price discrimination: Evidence from concession sales at movie theaters," *Working paper*, 2008.

[57] Grabiner, D., "Descartes' rule of signs: Another construction," *American Mathematical Monthly*, vol. 106, pp. 854–855, 1999.

[58] Greckel, F. and Felton, M. V., "Price and income elasticities of demand: A case study of Louisville," *Economic Efficiency and the Performing Arts*, pp. 62–73, 1987.

[59] Green, L., Kolesar, P., and Whitt, W., "Coping with time-varying demand when setting staffing requirements for a service system," *Production and Operations Management*, vol. 16, no. 1, pp. 13–39, 2007.

[60] Gross, D. and Harris, C., *Fundamentals of queueing theory.* John Wliey and Sons, 1985.

[61] Gurnani, H. and Tang, C., "Optimal ordering decisions with uncertain cost and demand forecast updating," *Management Science*, vol. 45, pp. 1456–1462, 1999.

[62] Gurumurthi, S. and Benjaafar, S., "Modeling and analysis of flexible queueing systems," *Naval Research Logistics*, vol. 51, pp. 755–782, 2004.

[63] Ha, A., "Optimal pricing that coordinates queues with customer-chosen service requirements," *Management Science*, vol. 47, pp. 915–930, 2001.

[64] Harel, A., "Convexity properties of the Erlang loss formula.," *Operations Research*, vol. 38, pp. 499–505, 1990.

[65] Harpaz, G., Lee, W., and Winkler, R., "Learning, experimentation, and the optimal output decisions of a competitive firm," *Management Science*, vol. 28, pp. 589–603, 1982.

[66] Harrison, J. and Lopez, M., "Heavy traffic resource pooling in parallel-server systems," *Queueing Systems*, vol. 33, pp. 339–368, 1999.

[67] Huntington, P., "Pricing policy and box office revenue.," *Journal of Cultural Economics*, vol. 17, no. 1, pp. 71–87, 1993.

[68] Jagerman, D., "Some properties of the Erlang loss function," *Bell System Technical Journal*, vol. 53, pp. 525–551, 1974.

[69] Jordan, W., Inman, R., and Blumenfeld, D., "Chained cross-training of workers for robust performance," *IIE Transactions*, vol. 36, pp. 953–967, 2004.

[70] Jorgensen, S., Kort, P., and Zaccour, G., "Optimal pricing and advertising policies for an entertainment event," *Journal of Economic Dynamics and Control*, vol. 33, pp. 583–596, 2009.

[71] Jouini, O. and Dallery, Y., "Estimating and announcing waiting times in multiple customer class call centers," *Proceedings of INCOM*, vol. 2, pp. 371–376, 2006.

[72] Jouini, O., Dallery, Y., and Nait-Abdallah, R., "Analysis of the impact of team-based organizations in call center management," *Management Science*, vol. 54, pp. 400–414, 2008.

[73] Kannan, P. and Kopalle, P., "To err is rational : Dynamic pricing on the internet: Importance and implications for consumer behaviour.," *International Journal of Electronic Commerce*, vol. 5, no. 3, pp. 63–83, 2001.

[74] Kelejian, H. H. and Lawrence, W. J., "Estimating the demand for Broadway theater: A preliminary inquiry," *Economic Policy for the Arts*, pp. 333–346, 1980.

[75] Kessler, M. and Turner, D., "Effects of density and cage size on stress in domestic cats (felis silvestris catus) housed in animal shelters and boarding catteries," *Animal Welfare*, vol. 8, no. 3, pp. 259–267, 1999.

[76] Ketchum, P., "Adopt-a-cat not an animal shelter," *Canadian Family Physician*, vol. 49, pp. 936–936, 2003.

[77] Kincaid, W. M., D. D., "An inventory pricing problem," *Journal of Mathematical Analysis and Applications*, vol. 7, pp. 183–208, 1963.

[78] Kochel, P., "Finite queueing systems structural investigations and optimal design," *International Journal of Production Economics*, vol. 88, pp. 157–171, 2004.

[79] Koole, G. and Pot, A., "An overview of routing and staffing algorithms in multi-skill customer contact centers: Technical report," *Department of Mathematics, Vrije Universiteit Amsterdam, The Netherlands*, 2006.

[80] Krishnan, K., "The convexity of loss rate in an Erlang loss system and sojourn in an Erlang delay system with respect to arrival rate and service rate," *IEEE Transactions on Communications*, vol. 38, no. 9, pp. 1314–1316, 1990.

[81] LA Animal Services, 2010. Official Website of Los Angeles Animal Services Department,
http://www.laanimalservices.com, last accessed on January 21, 2010.

[82] Ladany, S. and Arbel, A., "Optimal cruise-liner passenger cabin pricing policy," *European Journal of Operational Research*, vol. 55, pp. 136–147, 1991.

[83] Lan, Y., Gao, H., Ball, M., and Karaesmen, I., "Revenue management with limited demand information," *Management Science*, vol. 54, pp. 1594–1609.

[84] Lariviere, M. and Porteus, E., "Stalking information: Bayesian inventory management with unobserved lost sales," *Management Science*, vol. 45, pp. 346–363, 1999.

[85] Larsen, C., "Investigating sensitivity and the impact of information on pricing decisions in an $M/M/1/\infty$ queueing model," *International Journal of Production Economics*, vol. 56-57, pp. 365–377, 1998.

[86] Lazear, E., "Retail pricing and clearance sales," *American Economic Review*, vol. 76, pp. 14–32, 1986.

[87] Leloup, B. and Deveaux, L., "Dynamic pricing on the internet: Theory and simulations," *Journal of Electronic Commerce Research*, vol. 1, no. 3, pp. 265–276, 2001.

[88] Leslie, P., "A structural econometric analysis of price discrimination in Broadway theater," *Mimeo, Yale University.*, 1998.

[89] Leslie, P., "Price discrimination in Broadway theater," *RAND Journal of Economics*, vol. 35, pp. 520–541, 2004.

[90] Levin, Y., McGill, J., and Nediak, M., "Dynamic pricing in the presence of strategic consumers and oligopolistic competition," *Management Science*, vol. 55, pp. 32–46, 2009.

[91] Lieberman, W., "Implementing yield management," *ORSA/TIMS National Meeting, (San Francisco, California)*, 1992.

[92] Lin, K., "Dynamic pricing with real-time demand learning," *European Journal of Operations Research*, vol. 174, pp. 522–538, 2006.

[93] Little, J., "A proof of the theorem $L = \lambda W$," *Operations Research*, vol. 9, pp. 383–387, 1961.

[94] Littlewood, K., "Forecasting and control of passenger bookings." AGIFORS Symposium Proceedings 95-117, 1972.

[95] Lott, J. and Russell, R., "A guide to the pitfalls of identifying price discrimination," *Economic Inquiry*, vol. 29, no. 1, pp. 14–23, 1991.

[96] Lovejoy, W., "Myopic policies for some inventory models with uncertain demand distributions," *Management Science*, vol. 36, pp. 724–738, 1990.

[97] Maddiesfund Org., 2009. Official Website of Maddiesfund Organization, http://www.maddiesfund.org, last access on June 10, 2009.

[98] Maglaras, C. and Meissner, J., "Dynamic pricing strategies for multiproduct revenue management problems," *Manufacturing and Service Operations Management*, vol. 8, no. 2, pp. 136–148, 2006.

[99] MANDELBAUM, A. and REIMAN, M., "On pooling in queueing networks," *Management Science*, vol. 44, pp. 971–981, 1998.

[100] MARBURGER, D., "Optimal ticket pricing for performance goods," *Managerial and Decision Economics*, vol. 18, pp. 375–381, 1997.

[101] MASSEY, W. and WALLACE, R., "An optimal design of the M/M/C/K queue for call centers," *Queueing Systems*, vol. forthcoming, 2006.

[102] MATSUO, H., "A stochastic sequencing problem for style goods with forecast revisions and hierarchical structure," *Management Science*, vol. 36, pp. 332–247, 1990.

[103] MCKENNA, K., 2007. http://mckatie.wordpress.com, last accessed on January 12, 2010.

[104] MENDELSON, H., "Pricing computer services: queueing effects," *Communications of the ACM*, vol. 28, pp. 312–321, 1985.

[105] MESSERLI, E., "Proof of a convexity property of the Erlang B formula," *Bell system Technical Journal*, vol. 51, pp. 951–953, 1971.

[106] MOORE, T. G., "The demand for Broadway theater tickets," *Review of Economics and Statistics*, vol. 48, pp. 79–87, 1966.

[107] MOORE, T. G., "Economics of American theater," *Durham, N.C.: Duke University Press*, 1968.

[108] MURRAY, G. and SILVER, E., "A Bayesian analysis of the style goods inventory problem," *Management Science*, vol. 12, pp. 785–797, 1966.

[109] NAICS, 2009. Official Website of North American Industry Classification System (NAICS),
http://www.census.gov/naics, last access on October 11, 2009.

[110] NCPPSP, 2009. National Council on Pet Population Study and Policy (NCPPSP),
http://www.petfinder.com/for-shelters/facts-about-animal-sheltering.html, last access on June 29, 2009.

[111] NG, I., *The Pricing and Revenue Management of Services: A Strategic Approach.* Published By Routledge, An Imprint Of Taylor And Francis, under the Advances in Business and Management Studies, 2007.

[112] NO KILL ADVOCACY CENTER, 2009. Official Website of No Kill Advocacy Center,
http://www.nokilladvocacycenter.org, last access on March 3, 2009.

[113] Normando, S., Stefanini, C., Meers, L., Adamelli, S., Coultis, D., and Bono, G., "Some factors influencing adoption of sheltered dogs," *Anthrozoos*, vol. 19, no. 3, pp. 211–224, 2006.

[114] Ormeci, E., "Dynamic admission control in a call center with one shared and two dedicated service facilities," *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1157–1161, 2004.

[115] Palaka, K., Erlebacher, S., and D.H., K., "Lead-time setting, capacity utilization and pricing decisions under lead-time dependent demand," *IIE Transactions*, vol. 30, pp. 151–163, 1998.

[116] Petruzzi, N. and Dada, M., "Dynamic pricing and inventory control with learning," *Naval Research Logistics*, vol. 49, pp. 303–325, 2002.

[117] Plackett, R., "Karl Pearson and the Chi-Squared Test," *International Statistical Review*, vol. 51, pp. 59–72, 1983.

[118] Posage, J. M., Bartlett, P., and Thomas, D., "Determining factors for successful adoption of dogs from an animal shelter," *Journal of the American Veterinary Medical Association*, vol. 213, no. 4, pp. 478–482, 1998.

[119] PricewaterhouseCoopers, 2010. Official Website of PricewaterhouseCoopers,
http://http://www.pwc.com/gx/en/index.jhtml?ld=no, last accessed on October 4, 2010.

[120] Rascher, D., "A test of the optimal positive production network externality in major league baseball. in J. Fizel, E. Gustafson, and L. Hadley (Eds.) in Sports economics: Current research (pp. 27–45)," *Westport, CT: Greenwood*, 1999.

[121] Ray, S. and Jewkes, E., "Customer lead time management when both demand and price are lead time sensitive," *European Journal of Operational Research*, vol. 153, pp. 769–781, 2004.

[122] Reyniers, D., "A high-low search algorithm for a newsboy problem with delayed information feedback," *Operations Research*, vol. 38, pp. 838–846, 1990.

[123] Rosen, S. and Rosenfield, A., "Ticket pricing," *Journal of Law and Economics*, vol. 40, no. 2, pp. 351–376, 1997.

[124] Ross, S., *Stochastic processes.* John Wliey and Sons, 1996.

[125] Rothstein, M., "An airline overbooking model," *Transportation Science*, vol. 5, pp. 180–192, 1971.

[126] Scarf, H., "Bayes solutions to the statistical inventory problem," *Annals of Mathematical Statistics*, vol. 30, p. 1959, 490-508.

[127] Sen, A. and Zhang, A., "Style goods pricing with demand learning," *European Journal of Operation Research*, vol. 196, pp. 1058–1075, 2009.

[128] Sharma, P. and Singh, K., "Effect of shelter system on environmental variables," *Indian Journal of Animal Sciences*, vol. 72, no. 9, pp. 806–809, 2002.

[129] Sheikhzadeh, M., Benjaafar, S., and Gupta, D., "Machine sharing in manufacturing systems: Flexibility versus chaining," *International Journal of Flexible Manufacturing Systems*, vol. 10, pp. 351–378, 1998.

[130] Shumsky, R., "Approximation and analysis of a call center with flexible and specialized servers," *OR Spectrum*, vol. 26, pp. 307–330, 2004.

[131] Smith, D. and Whitt, W., "Resource sharing for efficiency in traffic systems," *Bell System Technical Journal*, vol. 60, pp. 39–55, 1981.

[132] SNKC Org., 2009. Spay Neuter Kansas City Organization (SNKC), http://www.snkc.net/petoverpopulation.html, last access on June 15, 2009.

[133] So, K. and Song, J., "Price, delivery time guarantees and capacity selection," *European Journal of Operational Research*, vol. 111, pp. 28–49, 1998.

[134] Steiner, F., "Optimal pricing of museum admission," *Journal of Cultural Economics*, vol. 21, no. 4, pp. 307–333, 1997.

[135] Subrahmanyan, S. and Shoemaker, R., "Developing optimal pricing and inventory policies of retailers who face uncertain demand," *Journal of Retailing*, vol. 72, pp. 7–30, 1996.

[136] Syski, R., *Introduction to Congestion Theory in Telephone Systems.* New York: North-Holland, 1986.

[137] Tadj, L. and Choudhury, G., "Optimal design and control of queues," *Spanish Statistical and Operations Research Society*, vol. 13, pp. 359–414, 2005.

[138] Toma, M. and Meads, H., "Recent evidence on the determinants of concert attendance for mid-size symphonies," *Journal of Economics and Finance*, vol. 31, no. 3, pp. 412–421, 2007.

[139] Travel and Leisure, 2010. Travel and Leisure Website, http://www.travelandleisure.com/articles/domestic-airlines-the-cost-of-pre-assigned-coach-seats/1, last accessed on September 20, 2010.

[140] Tseng, P., *Effects of Performance Schedules on Event Ticket Sales.* PhD thesis, University of Maryland, 2009.

[141] Wells, D., Graham, L., and Hepper, P., "The influence of length of time in a rescue shelter on the behaviour of kennelled dogs," *Animal Welfare*, vol. 11, no. 3, pp. 317–325, 2002.

[142] Wells, D. and Hepper, P., "The influence of environmental change on the behaviour of sheltered dogs," *Applied Animal Behaviour Science*, vol. 68, no. 2, pp. 151–162, 2000.

[143] Whitt, W., "Improving service by informing customers about anticipated delays," *Management Science*, vol. 45, pp. 192–207, 1999.

[144] Widder, D., *Advanced Calculus*. Dover Publications, Inc., New York, 1989.

[145] Williams, A., "Do anti-ticket scalping laws make a difference?," *Managerial and Decision Economics*, vol. 15, no. 5, pp. 503–509, 1994.

[146] Wolff, R. and Wang, C., "On the convexity of loss probabilities," *Applied Probability*, vol. 39, no. 2, pp. 402–406, 2002.

[147] Yakici, E., Duran, S., and Swann, J., "Dynamic switching times from season to single tickets with early switch to low demand tickets in sports and entertainment industry." Working Paper, Georgia Institute of Technology, 2009.

[148] Zawistowski, S.L., 2009. Petfinder Website, http://www.petfinder.com/for-shelters/euthanasia-statistics.html, last access on October 22, 2009.

[149] Zhao, W. and Zheng, Y., "Optimal dynamic pricing for perishable assets with nonhomogeneous demand," *Management Science*, vol. 46, no. 3, pp. 375–388, 2000.

[150] Ziya, S., Ayhan, H., and Foley, R., "Optimal prices for finite capacity queueing systems," *Operations Research Letters*, vol. 34, pp. 214–218, 2006.

VITA

Naragain Phumchusri was born in Bangkok, Thailand on January 1, 1983. She received her B.S. degree in Industrial Engineering from Chulalongkorn University, Bangkok, Thailand in 2004. After her graduation, she joined Procter & Gamble Manufacturing (Thailand) Ltd., where she worked as a production engineer with the responsibility of analyzing and reducing work-in-process inventory. She then pursued her graduate study and earned M.S. degree in Industrial and Systems Engineering at Georgia Institute of Technology in 2006. After that, she continued her doctoral study in the same department. In 2008, she worked as an intern at Predictix, a consulting company specializing on demand forecasting and promotion planning, where her deliverable was a statistical analysis identifying customers strategic behaviors affected by past promotions. Upon her Ph.D. graduation, she will be joining the Faculty of Engineering at Chulalongkorn University, Bangkok, Thailand.